

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 09:55:06 ; Search time 18166 Seconds

(without alignments)
6226.408 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985
Sequence: 1 MAGAPPALLPCLISDCC.....RSGVGPSTLPGSHRSPK 3104

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-O=/cgn2_1/USPTO.spool/US10016248/runat_18102004_105459_11541/app_query.faeta_1.3271
-DB=EST -QFMT=fastcap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10016248 @CGN 1.1_12209 @runat_18102004_105459_11541 -NCPU=6 -ICPU=3
-NO MMAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hc: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5729	33.7	6030	9	AY407073 Homo sapi
2	5696.5	33.5	5901	9	AY407075 Mus muscu
3	4805.5	28.3	3882	3	AK081081 Mus muscu
4	4001	23.6	5824	9	AY407074 Pan trogl
5	1284	7.6	748	7	CF744273 UI-M-GV0
6	1229.5	7.2	2841	3	AK046385 Mus muscu
7	1182	7.0	666	7	CF531622 UI-M-PY0
8	1126.5	6.6	2809	5	BO832212 C82 AF702
9	1122.5	6.6	673	5	BX671124 BX671124

10	1113.5	6.6	733	5	BU370222	603595935
11	1091	6.4	659	2	BB309178	BB309178
12	1053	6.2	707	7	CK312578	CK312578
13	1011	6.0	661	1	AL638364	AL638364
14	1000	5.9	656	6	BY734600	BY734600
15	933	5.5	671	5	BX676029	BX676029
16	931	5.4	3496	3	BC032550	BC032550
17	911.5	5.4	3348	3	AK003072	AK003072
18	894	5.3	666	4	RG080336	RG080336
19	885	5.2	686	7	CF742205	CF742205
20	884.5	5.2	803	4	BG217886	BG217886
21	872.5	5.1	3127	6	CB548140	CB548140
22	860.5	5.1	6075	3	AY261899	AY261899
23	854	5.0	3809	3	BC032548	BC032548
24	850	5.0	683	6	CD634832	CD634832
25	841.5	5.0	617	2	BF369877	BF369877
26	837.5	4.9	792	4	BG215218	BG215218
27	832.5	4.9	685	6	CD634834	CD634834
28	809	4.8	624	6	CD634836	CD634836
29	806	4.7	458	1	AT928999	AT928999
30	800	4.7	574	6	CD634867	CD634867
31	799	4.7	572	6	CD634862	CD634862
32	799	4.7	574	6	CD634863	CD634863
33	799	4.7	574	6	CD634869	CD634869
34	799	4.7	2888	9	AY414540	AY414540
35	796	4.7	4379	3	BC053923	BC053923
36	795	4.7	574	6	CD634865	CD634865
37	790	4.7	569	6	CD634866	CD634866
38	789.5	4.6	668	6	CD634821	CD634821
39	789.5	4.6	717	6	CD634825	CD634825
40	788	4.6	574	6	CD634859	CD634859
41	780	4.6	693	6	CD634827	CD634827
42	776.5	4.6	708	6	CD634837	CD634837
43	776.5	4.6	3663	3	AK083229	AK083229
44	773	4.6	568	6	CD634870	CD634870
45	772	4.5	3388	3	AK004995	AK004995

ALIGNMENTS

RESULT 1
AY407073
LOCUS
DEFINITION Homo sapiens CSMD1 gene, VIRTUAL TRANSCRIPT, partial sequence,
ACCESSION AY407073
VERSION AY407073.1 GI:39763044
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 6030)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sniensky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Infering nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL PUBMED
Science 302 (5652), 1960-1963 (2003)
14671302
REFERENCE 2 (bases 1 to 6030)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sniensky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
SOURCE location/Qualifiers
1..6030

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..26030
/gene="CSMD1"
/locus_tag="HCM2769"

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0	5729.00	6030	1008	259	505	59	1
Percent Similarity:	69.20%						
Best Local Similarity:	55.05%						
Query Match:	33.73%						

US-10-016-248-2 (1-3104) x AY407073 (1-6030)

```

QY      1251 LeuGluProGlyTyrAspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeu 1270
DB      1 ATGGAGCCCACTATGACTTCTTACACATCTATGAAGGGAAGATTCCAAACGCCCTC 60
QY      1271 IleGlySerPheTyrGlySerGlnLeuProGlyValGlyIleGluSerSerAsnSerLeu 1290
DB      61 ATTGGAGTTTACAGAGGCTCTCAGGCCCCAGAAAGAAATGAGAGTACGGAACAGCCTG 120
QY      1291 PheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThr 1310
DB      121 TTTCGGCATTTCCGAGTGTATGCTCCGTGGCCCTTCAAGGTTCCCATTTAA 180
QY      1311 GluAsnProArgGluSerCysPheAspProGlySerIleLysAsnGlyThrArgValGly 1330
DB      181 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCAGAAATATATGATGGGACAAAGATTGGA 240
QY      1331 SerAspLeuLysLeuGlySerSerValThrTyrCysHisGlyTyrGluValGlu 1350
DB      241 ACAGACTTCAGGCTTGCTGTCACCATCCATCCAGTGTGACTGTGGCTATTAAGATTCTT 300
QY      1351 GlyThrSerThrLeuSerCysIleLeuGlyProAspGlyLysProValITrpaAsnPro 1370
DB      301 GACCCCTATCATCATCACTGTGTGATGGGGCTGATGGAAACCTCTCGGACCAAGT 360
QY      1371 ArgProValCysThrAlaProCysGlyGlyGlnTyrValGlySerAspGlyValIleLeu 1390
DB      361 CTGCGCTTCCTGCAATGCTCCCTGTGAGGCGCAAGTACGCGGATCAGAAAGGGTATTTAA 420
QY      1391 SerProAsnTyrProGlnAsnTyrThrSerGlyGlnIleCysLeuTyrPheValIThrVal 1410
DB      421 TCACCAAACTAACCCCATTAATTACACAGCTGTCAAAATATGCTCTATTCATCCACGTA 480
QY      1411 ProLysAspTyrValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspVal 1430
DB      481 CCAAGAGATTCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGACACCTCGAATGATTGG 540
QY      1431 ValGluValHisAspGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySer 1450
DB      541 GCAGAAATTATTGATGGAACCCATGCACAGGCCACACTTCCAGCTCACTTCGGGGTCT 600
QY      1451 HisThrGlyGluSerLeuProLeuAlaThrSerAsnGlnValLeuIleLysPheSerAla 1470
DB      601 CACTCAGGGGAAACATTGCTTGGCTAGCTCAAAATCAATCTGCTCGATTGCACTGCA 660
QY      1471 LysGlyLeuAlaProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSer 1490
DB      661 AAGAGGGGTGCTCTCCCGCGGCTTCCACTTGGTATCAAGAGCTGTTCTCGTACCAAGT 720
QY      1491 AlaThrGlnCysSerSerValProGluProArgTyrGlyLysArgLeuGlySerAspPhe 1510
DB      721 GACACCAATGACAGCTCTGCTCCCGAGCCCAATGCGAAGAGAAATTGCTTCAAGTTT 780
QY      1511 SerValGlyAlaIleValArgPheGlyCysAsnSerGlyTyrAlaLeuGlnGlySerPro 1530
DB      781 TTTCGGGCTTCATGCTCGATTGCAATGCAACCCGGGATACCTGCTTCAAGGTTCCACG 840

```

```

QY      1531 GluIleGlyCysLeuProValITrpaGlyAlaLeuAlaGlnITrpaSerValSerAlaProThr 1550
DB      841 GCGCTCCACTGCGCAGTCCGTCGCCAACGCTTTGGCAGAGTGAACAGACGATCCCGCAG 900
QY      1551 CysValValProCysGlyGlyAsnLeuThrGluArgArgGlyThrIleLeuSerProGly 1570
DB      901 TGTGTGGTACCTGCGAGTGGCAATTTCACTCAACAGAAAGTTCATCTCTCCCGCG 960
QY      1571 PheProGluProTyrIleAsnSerLeuAsnCysValITrpaLysIleValValProGlyGly 1590
DB      961 TACCTGAGGCAATACGAAACAACTTGAACTGTATGGAAGATCATGATTACGAGCGCC 1020
QY      1591 AlaGlyIleGlnIleGlnValValSerPheValIThrGlnAsnITrpaAspSerLeuGlu 1610
DB      1021 TCGGAATTCAGATTCCAAGTGTATGCTTGGCAGAGACAGAACACTGGACTCCCTTAG 1080
QY      1611 ValPheAspGlyAlaAspAsnThrValIThrMetLeuGlySerPheSerGlyThrThrVal 1630
DB      1081 ATCCAGATGTGGGGATGTGACCCAGACCTGGGAAGCTTCTCAGNNNNNNNNNN 1140
QY      1631 ProAlaLeuLeuAsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSer 1650
DB      1141 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
QY      1651 ValSerAlaAlaGlyPheHisLeuGluTyrIleThrValGlyLeuSerCysProGlu 1670
DB      1201 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
QY      1671 ProAlaValProSerAsnGlyValIleLysThrGlyGluArgTyrLeuValAsnAspValVal 1690
DB      1261 CAGCCCTCCCGCAGCAAGACATCAAAATCGAAGATCGTCAATGAGTGAAGAGAGTGCTC 1320
QY      1691 SerPheGlnCysGluProGlyTyrAlaLeuGlnGlyHisAlaHisIleSerCysMetPro 1710
DB      1321 TCCTTCCAGTGGAGCGCGGGTACACCTCGAGGCGCGTCCCAATTTCTGTATGCCA 1380
QY      1711 GlyThrValAlaGlyITrpaAsnTyrProProProLeuCysIleAlaGlnCysGlyGlyThr 1730
DB      1381 GGGACCTGTCCGTTGGAACTATCCCTCTCCCTGTGCAATTCGACACTGTGAGGGAGC 1440
QY      1731 ValGluLeuMetGluGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsn 1750
DB      1441 CTGAGACCTTGGGTGTGTGATCTTCAAGCCCGGCTTCCAGATTTTCAACCCCAACAC 1500
QY      1751 MetAspCysSerITrpaLysIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeu 1770
DB      1501 TTAGACTGCACCTGGAGATCTCATTCACATCGGCTATGTGCACATATTCAGTTTCTG 1560
QY      1771 AsnPheSerThrGluProAsnHisAspTyrIleGluIleArgAsnGlyProTyrGlyThr 1790
DB      1561 AATTTTCTACCGAAGCTATATCAATGACTTCTTGAATAATTCAAATGACCTTCCACACC 1620
QY      1791 SerArgMetGlyArgPheSerGlySerGluLeuProSerSerLeuSerLeuSerITrpaSer 1810
DB      1621 AGCCCACTGATGACAAATTACCGGACAGGATCTCCCGCGCTCTGTGACACAAAG 1680
QY      1811 HisGluThrThrValTyrPheHisSerAspHisSerGlnAsnArgProGlyPheLysLeu 1830
DB      1681 CATGAACCTCTCACTCACTTATATGATGACATTCGCCAAACCGGACAGATTAAACTT 1740
QY      1831 GluTyrGlnAlaTyrGluLeuGlnGluCysProAspProGluProPheAlaAsnGlyIle 1850
DB      1741 GCTTACCAAGCCATTAATTAAGAACTGTCCAGATCCACCCCATTTAGAAATGGGTAC 1800
QY      1851 ValArgGlyAlaGlyTyrAsnValGlyGlnSerValIThrPheGluCysLeuProGlyTyr 1870
DB      1801 ATGATCAACTCGGATTAACGCTGGGGCAATCAGATATCTTGAAGTGTATCTGTGGTAC 1860
QY      1871 GlnLeuThrGlyHisProValLeuThrCysGlnHisGlyIThrAsnArgAsnITrpaHis 1890
DB      1861 ATTCTAATAGGCATCTGTCTCACTGTCAAGCATGGAGATCAACAGAAATCGGAATAC 1920

```

QY	1891	ProleuPcPolysCySgUValProCySgUlyGlyAenIleThrSerSerAanglyThrVal	1910
Db	1921	CCITTTCCAAAGATGATGCCCTTGTGGGATCAACGTAACTCTTCAGAACGGCACATC	1980
QY	1911	TySerProGlyIyPheProSerProTyzSerSerSerGlnAapCyValTrrLeuIleThr	1930
Db	1981	TACTCCCTGGCTTCTCGATGAGATGCCATCCGAAAGACATTCATTGGCTCATACG	2040
QY	1931	ValProIleGlyHiIleGlyValArgLeuAenLeuSerLeuLeuGlnThrGluProSergly	1950
Db	2041	GTGCCTCCAGGGCAGGAGATTATCAACACTTCAACCTGTTAACAAGCGAAGCTGTCAAC	2100
QY	1951	AapPheIleThrIleTrrAapGlyProGlnGlnThrAlaProArgLeuGlyValPheThr	1970
Db	2101	GATTACATGCTGTTGGAGCGGTCCCATCAAACTCAACCCAGCTGGAGTTTCAGT	2160
QY	1971	ArgSerMetAlaIyLeVLeThrValGlnSerSerSerAangInValLeuLeuIyPheHis	1990
Db	2151	GGCAACACAGCCCTCGAAACGGCGATAGCTCCACCAACCAAGTCTCTCAAGTTCCAC	2220
QY	1991	ArgAapAlaIaIaThrGlyIyIlePheAlaIleAlaPheSerAlaTyProLeuThrIys	2010
Db	2221	AGGCACTTTCCAAAGAGAGGCTTCTTGTCCTCAATTTCCAGCATTTCAAGTCAAGAA	2280
QY	2011	CySProPProProThrIleLeuProAaAlaGluValValThrGluAangGluPheAa	2030
Db	2281	TGTCAACTCCCTCCCGACGGGTCCACAGCAGAAAGCTTACTGAGGAATGAATTTCCAG	2340
QY	2031	IleGlyAapIleValArgTyrArgCySLeuProGlyPheThrLeuValGlyAaangIle	2050
Db	2341	ATAGGAGATTTTGTGATGATCCAGTGCACCCCGGGTACACCTGTGGGAGCGACAT	2400
QY	2051	LeuThrCySlySLeuGlyThrTyIleuGlnPheGlnGlyProProProIleCySgUVal	2070
Db	2401	CTGACTTCCAAGCTCAGTTCACAGTTGAGTTGAGGTTCTCTCCCAACTGTGAACA	2460
QY	2071	HiS CySProThrAaangGluLeuLeuThrAapSerThrGlyValIleLeuSerGlnSerTy	2090
Db	2461	CAATGCCCAGCAAAAGAAAGTCCGAGTCAATCATCGGAGATTCCTCAGTCCAGGGTAT	2520
QY	2091	ProGlySerTyProGlnPheGlnThrCySserTrrPLeuValArgValGluProAapTy	2110
Db	2521	CCGGGTAAATTTTAACTCCAGACTGTGCTTGGAGTAAATTAAGTGAACAAACATAC	2580
QY	2111	AaenIleSerLeuThrValGluTyPheLeuSerGlnIySgInTyAaPglPheGlnIle	2130
Db	2581	AACATTACATCTTGTGTGACACATTTCAAACTGAAGAAGCATTTGAAGCACTGGAACTG	2640
QY	2131	PheAapGlyProSerglyGlnSerProLeuLeuValaLeuSerglyAaenTySerAla	2150
Db	2641	TTTGTAGTTCCTTCTGGGCAAAAGTCTCTGCTAGAGCTTAAAGGGAAATCAATACGAA	2700
QY	2151	ProleuIleValThrSerSerSerAaenSerValTyIleAaGTPSerSerAapHiAla	2170
Db	2701	CAATCAAAATTTTACAGACAGAGATATAGTTAATCTCCGCTGGTCCATGACATGATGCC	2760
QY	2171	TyrAaenArglySgIyPheIySIIeArgTySerAlaProTyTyCySerLeuProArgAla	2190
Db	2761	ACCACTAAGAAAGATTCMAAGATTGCTATGACAGCACTTACAGCTTTAACCACCCC	2820
QY	2191	ProleuHiIleGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHiIlePhe	2210
Db	2821	CTGAAGAAATGGGGGATTCCTAAACAGCACTCAGAGAGCGGTTGGAAACCAAGATCATAT	2880
QY	2211	GlyCyAaenAlaGlyTyPArgLeuValGlyIyHisSerMetAlaIleCySThrArgHiPro	2230
Db	2881	TTTGGCAAGCCTGATACCGAATGGTGGCCACAGCAATGCAACTGTGAACGAACCCA	2940
QY	2231	GlnGlyTyTrIleLeuTrrSerglyAlaIleProleuCySgInAlaLeuSerCySgIyLeu	2250
Db	2941	CTTGGCAATGACAGTGGAGATCTCCCTACAGCACTCTCCAGCGCTGTCTCTGTGGATC	3000
QY	2251	ProGluAlaProIyAaangIyMeValPheGlyIySgUlyTyThrValGlyThrIyAla	2270

Db	3001	CCAGATCCCCAGGAACGGTTCAATTACCGGGAAACGATTCACCTTGAGACATGTAAGTGTG	3060
Qy	2271	ValTyrSerCysSerGlnGlyTyrPheGlnAlaGlyAlaGlnAlaThrAlaGluCys	2290
Db	3061	GTCATATGAATTCATGAGCGCTTCAAGCTTGAATTCACCGACCAAGCAACGCGGTGTG	3120
Qy	2291	LeuAspThrGlyLeuTyrTrpSerAsnArgAsnValProProGlnCysValProValThrCys	2310
Db	3121	CAGAAAGATGGCGCTGTGGAGATGACAAAGAGGGAAGCCGCCATGTGTAAAGCCGCTGCTTGC	3180
Qy	2311	ProAspValSerSerIleSerValGlnHisGlyArgTyrPargLeuIlePheGlnThrGln	2330
Db	3191	CCGAGCATTTGAGCTCAGCTCTCTCAGAAATATTCATCTGGAGCGTGGTTTCAGATCTTGG	3240
Qy	2331	TyrGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrTyrThrGlyGln	2350
Db	3241	AATAGATACGGGTGCTCAAGATATTCGTAGCTGACGTCTCTGTTACTTAAAGGCTGCG	3300
Qy	2351	ArgValIleArgCysGlnAlaAsnGlyLysTrpSerLeuGlyAspSerThrProThrCys	2370
Db	3301	AGGCTCTCGCGGTGCCAGGCCAAATGGGACGTGAAACATAGAGATGAGAGGCCAAAGCTGT	3360
Qy	2371	ArgIleIleSerCysGlyGluLeuProIleProProAsnGlnHisArgIleGlyThrLeu	2390
Db	3361	CGA-----	3363
Qy	2391	SerValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrLeuValGlySer	2410
Db	3363	-----	3365
Qy	2411	ArgValArgGluCysMetAlaAsnGlyLeuTrpSerGlySerGluValArgCysLeuAla	2430
Db	3364	-----GCT	3366
Qy	2431	GlyHisCysGlyThrProGluProIleValAsnGlyHisIleAsnGlyGluAsnTyrSer	2450
Db	3367	GGCCACCTCGGTTCCCCAGACCCTATGTGAACGCTCATTAATGGAGAGGCTTCAGT	3426
Qy	2451	TyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIleGlyMetSerVal	2470
Db	3427	TACAGAGACACGGGTGGTTTACCAGTGCATCTCGTTCGCGCTTGGAGAACTCCGTG	3486
Qy	2471	ArgIleCysGlnGlnAspHisAsnIleTrpSerGlyLysThrProPheCysValProIleThr	2490
Db	3487	AGGATATCCCTCGAAGACCAACAAGTGGTCTGACAAAGCCGTGTGTGCTCCATGACA	3546
Qy	2491	CysGlyHisPProGlyAsnProValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsn	2510
Db	3547	TGTGGTCACTCTGGAAACCTCGCCACGAGTATCTAATGACAGTGAATTCACCTGAT	3606
Qy	2511	AspValValysPheValCysAsnProGlyTyrMetAlaGluGlyAlaIleArgSerGln	2530
Db	3607	GATGTCGGAATTTCACCTGCAACACGGGCTATTGTCGACGGGCGGTGTCTCGAGCCAG	3666
Qy	2531	CysLeuAlaSerGlyGlnTrpSerAspMetLeuProThrCysArgIleIleAsnCysThr	2550
Db	3667	TGTGGAGCAACGGCGCAGTGGAGTAAAGCCCTCTGCCACAGTGTGCAAGTGTAACTGTCT	3726
Qy	2551	AspProGlyHisGlnGluAsnSerValArgGlnValHisAlaSerGlyProHisArgPhe	2570
Db	3727	GATCCAGGCTTTGTGGAAATGCGCATTCGTCTACCGGCAACAGAACTTCCCTCGAAGTTTT	3786
Qy	2571	SerPheGlyThrThrValSerTyrArgCysAsnHisGlyPheTyrLeuLeuGlyThrPro	2590
Db	3787	GAGTATGAATAGAGATCTCTGTACCATTTGCAGAAGGATTTTACTTGCTGGGATCTTCA	3846
Qy	2591	ValLeuSerCysGlnGlyAspGlyThrTrpAspArgProArgProGlnCysLeuLeuVal	2610
Db	3847	GCGTTGACCTGTAAAGCAAAATGGCTTAATGGGACCGATCCCGCCAAAGTGTGGCTATA	3906
Qy	2611	SerCysGlyHisPProGlySerProProHisSerGlnMetSerGlyAspSerTyrThrVal	2630

```

Db      3907 TCCTGTGACACCCAGGGGTCCCTGCCAACGCCGTCTCATGAGAGCTTTTACTAT 3966
Qy      2631 G1VALValValArgTyrSerCysIleGlyValArgThrLeuValGlyAsnSerThrArg 2650
Db      3967 GGGCGCGTGTGACACTTCTCTGTCAGAGGAGCGAGGCTTCATGAGCAACACACGAGA 4026
Qy      2651 MetCysGlyLeuAspGlyHisTyrThrGlySerLeuProHisCysSerGlyThrSerAl 2670
Db      4027 GTGTGTCAGAGAGACAGTCACTGAGCGGGGCACTGCCCACTGCACAGAAATATCTCT 4086
Qy      2671 G1VALValValAspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAsp 2690
Db      4087 GGAATCTGTGTGATCCGGGAGCCCGAGACATGGGCTCGGCTTGTGTGATGACTTTAG 4146
Qy      2691 ProGlyThrValMetArgPheSerCysGluAlaGlyHisValLeuArgGlySerSerGlu 2710
Db      4147 ACAAAAGATCTTCCGCTTCTCTGTGAATGGGGCCACGAGGGGGCTCCCTGAA 4206
Qy      2711 ArgThrCysGluAlaAsnGlySerTyrPserGlySerGlnProGluCysGlyValIleSer 2730
Db      4207 CGCAGGTGTGTGCTCAATGGGTGATGATGAGGACTGACGCGGTGTGTGAGCGGTCTCC 4266
Qy      2731 CysGlyAsnProGlyThrProSerAsnAlaArgValValPheSerAspGlyLeuValPhe 2750
Db      4267 TGTGCAACCTCTGGACACCAACAGGAATGATGTGACTGATGATGGCATTTCTGTC 4326
Qy      2751 SerSerSerIleValTyrGluCysArgGluGlyTyrTyrAlaThrGlyLeuLeuSerArg 2770
Db      4327 TCCAGCTGTGATCATCTATGCTGTGGGAGGCTACAAAGCCTGAGGCTCATGACACGG 4386
Qy      2771 HisCysSerValAsnGlyThrTyrThrGlySerAspProGluCysLeuValIleAsnCys 2790
Db      4387 CATTCGACAGCCCAATGGAGCTGAGCAGGCACTGCTCCCACTGCACAAATTAATGTTGT 4446
Qy      2791 GlyAspProGlyIleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgTyrAsnIle 2810
Db      4447 GGGATCCAGGCAACACTAGCAAAATGGGATCAGTTTGGACCACTTCACTTCAACAG 4506
Qy      2811 ThrValThrTyrGlnCysValProGlyTyrMetMetGlnSerHisArgValSerValLeu 2830
Db      4507 ACTGTGACTATCATGTATGATCCAGGCTATGTCATGAGAACATGACATCCGCTCATTT 4566
Qy      2831 SerCysThrTyrAspArgTyrThrPAsnGlyThrIleProValCysIleAlaMetCys 2850
Db      4567 CGCGTACCAAGACCGGAGAGTGAATCCAGCAACCTGTCTGCAAGCGGCTGTGT 4626
Qy      2851 LysProProProLeuIleProAsnGlyLysValValGlySerAspPheMetTyrGlySer 2870
Db      4627 CCTCAGCGCGCGCGGTGAGAAATGGAACAGTGAAGGAAATGATTTCCGCTGGGCTCC 4686
Qy      2871 SerValThrTyrAlaCysIleGluGlyTyrGlnLeuSerLeuProAlaValPheThrCys 2890
Db      4687 AGCATTAAGTACAGCTGATGAGCGGTACCACTCTTCACTCCGCCATCTCTCTGT 4746
Qy      2891 GlnGlyAsnGlySerTyrThrGlyLeuLeuProGlnCysPheProValPheCysGlyAsp 2910
Db      4747 GAAGGTGCGGGGTGTGGAAGAGAGATCCCACTGTCTCTGTGTCTTCTGCGGAGAC 4806
Qy      2911 ProGlyValProSerArgGlyArgGlyAspArgGlyPheSerTyrArgSerSerVal 2930
Db      4807 CCTGGCATCCCGCAAGAGGCACTTAGTGGAAAGTTTCACTATTAAGTCCGAAAGTC 4866
Qy      2931 SerPheSerCysHisProProLeuValLeuValGlySerProArgArgPheCysGlnSer 2950
Db      4867 TTCTTCAGTGAACATCTCATTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4926
Qy      2951 AspGlyThrTyrPserGlyThrGlnProSerCysIleAspProThrLeuThrCysAla 2970
Db      4927 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4986
Qy      2971 AspProGlyValProGlnPheGlyIleGlnAsnAsnSerGlnGlyTyrGlnValGlySer 2990
Db      4987 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5046

```

```

Qy      2991 ThrValLeuPheArgCysGlnIleGlyTyrLeuLeuGlnGlySerThrThrArgThrCys 3010
Db      5047 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5106
Qy      3011 LeuProAsnLeuThrTyrPserGlyThrProProAspCysValProHisCysArgGln 3030
Db      5107 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5166
Qy      3031 ProGlyThrProThrHisAlaAsnValGlyAlaLeuAspLeuProSerMetGlyTyrThr 3050
Db      5167 CCAGAAACCCCGGACACCGGATGTGAGAGCCATGATCTTCTTACTTTGCGCTACACC 5226
Qy      3051 LeuIle-ThrProAlaArgArgAlaSerProSerArgValAlaProSerThrAlaProAl 3070
Db      5227 TTAGTGTACACCTGCGCATTCAGCTTTTCTCTGCGAGGGGATCTGAGACAGAACATGT 5286
Qy      3070 AARGArgMetAlaAlaGlyGlnAlaSerArg 3080
Db      5287 AAAGACAGATGAATGACAGAAAGTCGC 5317

RESULT 2
AY407075      5901 bp      DNA      linear      GSS 15-DEC-2003
LOCUS      Mus musculus CSM1 gene, VIRUTAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY407075.1 GI:39763046
VERSION      AY407075.1 GI:39763046
KEYWORDS      GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 5901)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A.,
Todd,M.A., Tenenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Smirsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source      location/Qualifiers
gene      1..5901
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            <1..>5901
            /gene="CSM1"
            /locus_tag="HGM2769"

ORIGIN
Alignment Scores:
Pred. No.:      0      Length:      5901
Score:      5696.50      Matches:      1014
Percent Similarity:      69.02%      Conservative:      249
Best Local Similarity:      55.41%      Mismatches:      467
Query Match:      33.54%      Indels:      103
DB:      9      Gaps:      4

US-10-016-248-2 (1-3104) x AY407075 (1-5901)
Qy      1251 LeuGluProGlyTyrAspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeu 1270

```



```
Db      1 ATGAGCGCGAGTATGACTTCTCCATATCTATGAGGAGGAGACTCCAAACGCCACTG 60
Qy      116G1SerPheTyrglySerGlnLeuProGlyArgIleGluSerSerSerSerLeu 1230
Db      61 ATTGGAGTTTCCAGGGCTCTCAAGCCCAAGAGAGATCGAGAGCATGTGTAACGCTC 120
Qy      1291 PheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGlyPheValIleAspTyThr 1310
Db      121 TTTCGGGATTCAGAGAGTATGATCTTCTGTCGCTGCGGATTCGCCATTGAATTAA 180
Qy      1311 GlnAsnProAlaGlySerCysPheAspProGlySerIleLysAsnGlyThrArgValGly 1330
Db      181 GAGAAACCCCGGAGACTTGTCTTGACCCCTGGAAACAATGAATGAGACAAGATTGGA 240
Qy      1331 SerAspLeuLysLeuGlySerSerValThrTyThrCysHisGlyGlyThrGlnValGlu 1350
Db      241 ACTGACTTTAACTGAGGCTCTACAGTTACCTATCAGTGTGACTCTGTGTACAAAGATTGTG 300
Qy      1351 G1yThrSerThrLeuSerCysIleLeuGlyProAspGlyLysProValITPAsnAspPro 1370
Db      301 GATCCCTGCTCATTTAGTGTGTGACAGGGGCTGATGGGAAGCCGTCCTGSCA----- 353
Qy      1371 ArgProValCysThrAlaProCysGlyGlyGlyThrValGlySerAspGlyValIleu 1390
Db      354 -----CCCTGTGAGGCGAGTACAGGGGCTCGAGGGGGTGTGTTG 395
Qy      1391 SerProAsnTyrrProGlnAsnTyrrThrSerGlyGlnIleCysLeuTyrrPheValThrVal 1410
Db      396 TCACCAACTGACCTCATCTATCTACACAGTGGAGATGTGATCTATCTCATTCATCAGCTG 455
Qy      1411 ProLysAspTyrrValAlaPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspVal 1430
Db      456 CCCAAGGAATTCGTGTATTGTTGGACAGTTTCCCTATTCCAGACTGCATGAATGACTTG 515
Qy      1431 ValGluValHisAspGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySer 1450
Db      516 GCAGAGTTGTTGTGGAACACATCCCGACGCGAGCTTGT----- 556
Qy      1451 HisThrGlyGluSerLeuProLeuAlaThrSerAsnGlnValIleLysPheSerAla 1470
Db      557 -----GAAACACTGCGCACTGAGTACATCCATCAGATTCGCTTCGATTGAGCGCC 607
Qy      1471 LysGlyLeuAlaProAlaArgGlyPheHisPheValTyrrGlnAlaValProArgThrSer 1490
Db      608 AAAAGCGAGCTTCTGCAAGGGGTTTCCACTTCTCTCAAGCTGTTCCGCGACCACT 667
Qy      1491 AlaThrGlnCysSerSerValProGluProArgTyrrGlyLysArgLeuGlySerAspPhe 1510
Db      668 GATACACAGTGCAGCTGTCCCGAGCCAGATATGGAGAGAAGATTGTTGAGATTTC 727
Qy      1511 SerValGlyAlaIleValArgPheGlyCysAsnSerGlyTyrrAlaLeuGlnGlySerPro 1530
Db      728 TCCCAGAGCTCATCTGCTCGCTGAGTGAACCCAGGTTACTCTGTCGCAAGGTTCCACC 787
Qy      1531 GluIleGlyCysLeuProValProGlyValaLeuAlaGlnITPAsnValSerAlaProThr 1550
Db      788 GCCATCCGCTCCAGCTCTGCTGCCAAATGCTTTGGCTCAGTGAATGACACATCCCAAGC 847
Qy      1551 CysValValProCysGlyGlyAsnLeuThrGlnArgArgGlyTyrrIleLeuSerProGly 1570
Db      848 TGTTAGTTCATGACAGTGCAGCACTTCACTCAGAGAAAGAGGAGATTGTTTCCCGAGGC 907
Qy      1571 PheProGluProTyrrLeuAsnSerLeuAsnCysValITPlyValIleValValProGluGly 1590
Db      908 TATCTTGAGCCCTTGGGACACATGAATCTGTGTGGAAGATCATGTATTCGAGAGGC 967
Qy      1591 AlaGlyIleGlnIleGlnValValSerPheValThrGluGlnAsnITPAsnSerLeuGlu 1610
Db      968 TCGGGATTCAGATTCAGATGATGACTTTGCCACAGACAACTGGGACTCCCTGGAG 1027
Qy      1611 ValPheAspGlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrVal 1630
Db      1028 ATCCATGATGAGAGACATGACGCGCCGAGACTGCGGACGCTTCTAGNNNNNNNNNN 1087
Qy      1631 ProAlaLeuLeuAsnSerThrSerAsnGlnLeuTyrrLeuHisPheTyrrSerAspIleSer 1650
Db      1088 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1147
Qy      1651 ValSerAlaAlaGlyPheHisLeuGluTyrrLysThrValGlyLeuSerSerCysProGlu 1670
Db      1148 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1207
Qy      1671 ProAlaValProSerAsnGlyValLysThrGlyGlnArgTyrrLeuValaAspValVal 1690
Db      1208 CTTGCTCTCCCGACCAATGCGATCAAGATGAGAACCGGTACAGTGAACGATGTCCTG 1267
Qy      1691 SerPheGlnCysGluProGlyTyrrAlaLeuGlnGlyHisIleAsnHisIleSerCysMetPro 1710
Db      1268 TCTTTCAGTGTGAACCTGGGTATACCTTGCAGGGCCGCTCAGACATTTCTTGATGCCA 1327
Qy      1711 G1yThrValArgArgITPAsnTyrrProProLeuGlyCysIleAlaGlnCysGlyGlyThr 1730
Db      1328 GGAACGTAAGTGTGGAACCTACCTCCCTTGTGTGATGCTCACCTGTGGTGGGACA 1387
Qy      1731 ValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAsnTyrrProSerAsn 1750
Db      1388 CTAACCAAGATGATGAGATGATCTTGAGCGCCAGGGTTCACAGGCTCATACCCCAACAC 1447
Qy      1751 MetAspCysSerITPlyHisIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeu 1770
Db      1448 TTGAGCTGACCTGGAGATATACCTGCCATTTGGATGTGACATATCCAAATTTCTA 1507
Qy      1771 AsnPheSerThrGluProAsnHisAspTyrrIleGluIleArgAsnGlyProTyrrGluThr 1790
Db      1508 AATTTCATCTGAAGCAACACATGACTACTGAGATCCAGAAATGAGCCCTTACACAGC 1567
Qy      1791 SerArgMetCecGlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSer 1810
Db      1568 AGTCCAAATGATGGGCCAGTTCAAGTGGCCCTGACTCATATGCTGAGACACACA 1627
Qy      1811 HisGluThrThrValTyrrPheHisSerAspHisSerGlnAsnAspProGlyPheLysLeu 1830
Db      1628 CACGAAACCTCATCCCTCTATATGATGACACTACAGAAACCGGCAAGGATTTAACTC 1687
Qy      1831 GluTyrrGlnAlaTyrrGluLeuGlnGlyCysProAspProGluProPheAlaAsnGlyIle 1850
Db      1688 AGTTACCAAGTTATGATGATTACAGAACTGCCAGATCCACCCGATTCAGAAATGGGTTTC 1747
Qy      1851 ValArgGlyAlaGlyTyrrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTyrr 1870
Db      1748 ATGATCAACTCGGATTTACAGCGTGGGCCAGATCGCATCTCTTTGAGTGTCTATCCCGGCTAC 1807
Qy      1871 GlnLeuThrGlyHisPheProValLeuThrCysGlnHisGlyThrAsnArgAsnITPAspHis 1890
Db      1808 ATCTTGTTAGGCCACCTGTGCTCACTGCGCAGATGAACCTGACAAACTGGAACCTAC 1867
Qy      1891 ProLeuProLysCysGlyValProCysGlyGlyAsnIleThrSerSerAsnGlyThrVal 1910
Db      1868 CTTTCCACAGGTATGACGCTCCTGTGATATATGTGATATGACATACAGATGTGACATT 1927
Qy      1911 TyrrSerProGlyPheProSerProTyrrSerSerSerGlnAspCysValITPLeuIleThr 1930
Db      1928 TATCCCTCGGGTTCCAGATGAGTACCAATTCAGTAAGAGACTGTCTGTGGCTGTGACT 1987
Qy      1931 ValProIleGlyHisGlyValaArgLeuAsnLeuSerLeuLeuGlnITPThrGluProSerGly 1950
Db      1988 GTCCCTCAGGCCACGAGGTGATCAATTAATTCCTTGTCGACAGCCGAGGCTGTAAAT 2047
Qy      1951 AspPheIleThrIleITPAspGlyProGlnGlnITPAspArgLeuGlyValaPheThr 1970
Db      2048 GACTACATTTGCTGTGTGATGTGCTGACCAAGAACTCGCTAGCTCGGGGTTTCACT 2107
Qy      1971 ArgSerMetAlaLysTyrrValGlnSerSerSerAsnGlnValaLeuLeuLysPheHis 1990
Db      2108 GGAACACTGCGCTTAGACAGCTTACAGCTCCACCAACAGAGGCTGTCAATTCAC 2167
```

Qy 1991 ArgAspAlaAlaThrGlyGlyIlePheAlaIleAlaPheSerAlaTyrProLeuThrIle 2010
Db 2168 AGGATTTCTTCATGAGAGCTTCTTCTCTCAATTTCCATGATTTCACTGGAAGG 2227
Qy 2011 CysProProProThrIleLeuProAsnIleValValIleThrGluAsnGluPheAsn 2030
Db 2228 TGCCCGCTCCCGCAGAGTGCACAGGCTGACTGCTTACAGAAAGTGAAGACTTTGAA 2287
Qy 2031 IleGlyAspIleValAlaGlyTyrArgCysLeuProGlyPheThrIleValGlyAsnGluIle 2050
Db 2288 ATAGGGAGCTTCGGAAGTACCAAGTCCATCCAGGGTACCAATTTGGGAGACGACCC 2347
Qy 2051 LeuThrCysIleuGlyThrTyrLeuGluIlePheGluIleProProIleCysGlyVal 2070
Db 2348 CTGACATGCAAGCTCAGCTCAGACTACTGTTCCAGAGCTCTCCACCCACTGTGAGCA 2407
Qy 2071 HisCysProThrAsnGluLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTyr 2090
Db 2408 CAATGCCCAATATGAAAGTGCAGACAGCTTCTGGGTGATTTCTCACTCCCGGGTAC 2467
Qy 2091 ProGlySerTyrProGlnIlePheGlnThrCysSerTyrLeuValArgValGluProAspTyr 2110
Db 2468 CCAAGGCAACTATTTAACTCCAGACATGCTTGAGTAAAGTGAAGCCAACTTC 2527
Qy 2111 AsnIleSerLeuThrValGlyTyrPheLeuSerGlyIleGlyIleThrAspGluPheGluIle 2130
Db 2528 AACATCACTCTTTGGTGCACACTTCCAAAGTAAAGCAATTTGATGACGTGAAGT 2587
Qy 2131 PheAspGlyProSerGlyIleSerProLeuIleValAlaLeuSerGlyAsnTyrSerAla 2150
Db 2588 TTTGATGTTCTTCTGACAAAGCTCTGTAGTGTCTTAAGTGGAAACCACTGAA 2647
Qy 2151 ProLeuIleValThrSerSerSerSerValTyrLeuArgTyrPheSerAspIleAla 2170
Db 2648 CAGTCAATTTTACAGACAGAACATGACATCTGCTGCTGCTCAAGACCAATGCA 2707
Qy 2171 TyrAsnArgIleGlyPheIleValIleArgTyrSerAlaProTyrCysSerLeuProAlaGly 2190
Db 2708 ACCAGCAAGAAAGATTCAAGATTGCTATGCAAGCTCTTACTGCAAGCTCACTCTACA 2767
Qy 2191 ProLeuIleGlyPheIleLeuGlyIleThrSerThrGlnProGlyIleSerIleHisPhe 2210
Db 2768 CTCAGCAAGTGGCATTTTAAATAAACCGCTGGGCGCACTGGAGCAAGTAACTAC 2827
Qy 2211 GlyCysAsnAlaGlyTyrArgLeuValGlyHisSerMetAlaIleCysThrArgHisPro 2230
Db 2828 TTCTGCAAGCTGATATGCAATGATTGGCCACAGCAATGCACTGCAAGCGAAACCCA 2887
Qy 2231 GlnGlyTyrHisLeuTyrPheSerGlnAlaIleProLeuCysGlnAlaLeuSerCysGlyLeu 2250
Db 2888 GTGGGCTGTATAGTGGAGCTCGATGGACCACTTTGGCAAGCTGTGCTTGTGAATT 2947
Qy 2251 ProGlnAlaProIleValAsnGlyMetValPheGlyIleGlyIleThrValGlyThrIleVal 2270
Db 2948 CCAAGAGCTCTCGGAAATGCTCATTCACCGGCAATGAGTTCACTTACAGCAATAAAGT 3007
Qy 2271 ValTyrSerCysSerGlyIleGlyTyrHisIleGlnAlaGlyValAlaIleThrAlaGluCys 2290
Db 3008 ACTTATGATGTATATGAAGCTTCAAGCTGATGCGATGCAAGACCACTGTATGT 3067
Qy 2291 LeuAspThrGlyLeuTyrPheSerAsnArgAsnValProProGlnCysValProValThrCys 2310
Db 3068 CAGAAGATGCTGTGTCAGCAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3127
Qy 2311 ProAspValSerSerIleSerValGluHisGlyArgTyrArgLeuIlePheGluThrGln 2330
Db 3128 CCCAGAGCTGCAAGCGCAAGCTGTGAGAGCAAGCTGTGAGAGCTGTGAGAGCTGTG 3187
Qy 2331 TyrGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGln 2350
Db 3188 AATGAATACGAGACTCAAGTTCTCTCAAGCTGCAAGCCCGGCTTCTTGGAGGGTACG 3247

Qy 2351 ArgValIleArgCysGlnAlaAsnGlyIleTyrPheSerLeuGlyAspSerThrProThrCys 2370
Db 3248 AGGCTTTGACGTGTCAAGCCATGGAGACTGGAGACTGGAGACTGGAGAAACAGACC 3301
Qy 2371 ArgIleIleSerCysGlyIleLeuProIleProProAsnGlyHisArgIleGlyThrLeu 2390
Db 3301 3301
Qy 2391 SerValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrIleValGlySer 2410
Db 3301 3301
Qy 2411 ArgValArgGluCysMetAlaAsnGlyLeuTyrPheSerGlySerGluValArgCysLeuAla 2430
Db 3302 3302
Qy 2431 GlyHisCysGlyThrProGluProIleValAlaAsnGlyHisIleAsnGlyIleAsnTyrSer 2450
Db 3314 GGTCAATTGGCTCTCCAGACCCCAATTTGAAATGGCCATATACGCGGTATGATTTCAAC 3373
Qy 2451 TyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIleGlyMetSerVal 2470
Db 3374 TACAGAGACAGAGTGTCTTACCAATGCAACCTGGGTTCGGCTGTGGGCACTTGTG 3433
Qy 2471 ArgIleCysGlnGlnAspHisIleTyrPheSerGlyIleThrProPheCysValProIleThr 2490
Db 3434 AGAATTTGCTGTACAGACCAAGTGTGTGGGAGAGCTCTGTTGGCTCCCAATCA 3493
Qy 2491 CysGlyHisProGlyIleAsnProValAsnGlyIleThrGlnGlyAsnGlnPheAsnLeuAsn 2510
Db 3494 TGTGACACCTCGGAAACCTTCGATGCTTCCACCAACGGGACCGGAGTTCAACCTGAAC 3553
Qy 2511 AspValIleValPheValCysAsnProGlyTyrMetAlaGluGlyValAlaIleAsnSerGln 2530
Db 3554 GACCTTGTAAATTTCACTGCTCCATCTGCTACCGGCTGCAAGGCTCTCCGAGGCCAG 3613
Qy 2531 CysLeuAlaSerGlyGlnTyrPheSerAspMetLeuProThrCysArgIleIleAsnCysThr 2550
Db 3614 TGTGGAGCAAGGCGCAAGTGGAGCAAGCCCTTGGCCATGTGCGAGTGTGAATCTTCC 3673
Qy 2551 AspProGlyIleGlnGluAsnSerValArgGlnValHisAlaSerGlyProHisArgPhe 2570
Db 3674 GACCCCGGATCTGTGAAATGACATGCTCCGATGGGCAACAGAACTTCCACAGAGTTTC 3733
Qy 2571 SerPheGlyThrThrValSerTyrArgCysAsnHisIleGlyPheTyrLeuGlyIleThrPro 2590
Db 3734 GAGTATGGGACAGTGTGTATGATCACTGCAAAACGGGGTCTTACTGTGGGATTTCT 3793
Qy 2591 ValLeuSerCysGlnGlyAspGlyThrTyrPheAspArgProArgProGlnCysLeuLeuVal 2610
Db 3794 GCCCTGACCTGATGGCAAGTGGCTTTGGGAGCCGCTCTTACCAAGTGTCTGGCTATA 3853
Qy 2611 SerCysGlyHisProGlySerProProHisSerGlnMetSerGlyAspSerTyrThrVal 2630
Db 3854 TCATGTGGGATCTCGGGGTCTCTCTATAGTCTGCTGCACTGAGAAATGTTTACATAC 3913
Qy 2631 GlyAlaValValArgTyrSerCysIleGlyIleValGlyThrIleValGlyAsnSerThrArg 2650
Db 3914 GAGGCAAGTTCAGATCTCTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3973
Qy 2651 MetCysGlyLeuAspGlyHisIleTyrThrGlySerLeuProHisCysSerGlyThrSerVal 2670
Db 3974 GTCTGCAAGAAAGACAGTACAGTGGATGATCTCTTCCATTTGTTCAAGAAATGTCT 4033
Qy 2671 GlyValCysGlyAspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAsp 2690
Db 4034 GGATTTTGGTGTGATCCAGGAGCCCAAGACATGAGTCTGCTTGGGAGTATGATTAAAG 4093
Qy 2691 ProGlyIleValMetArgPheSerCysGlnAlaGlyHisValLeuArgGlySerSerGlu 2710
Db 4094 ACAAAAGCTTTTCCGCTTCTCTGTGATGAGTGGGCAACAGCTCGGGGCTCTGCAAGA 4153
Qy 2711 ArgThrCysGlnAlaAsnGlySerTyrPheSerGlySerGlnProGluCysGlyValIleSer 2730

```

Db      4154 CGCAGATGCTGTGAATGATCTCTGGTCAGAGAGACGCTGTGTGAGCCGCTGCC 4213
Qy      2731 Cy6el1yAenProgl1yThrProserAenAlaArgValAlaPheSerAepGlyLeuValPhe 2750
Db      4214 TGCAGAAACCCAGACGCGCCAGCCAGATGATCTCCAGACATGTATGGATCTCTTC 4273
Qy      2751 SerSerSer1Leu1yThrGluCyAArgGluGlyTyrrTyrrAlaThrGlyLeuLeuSerArg 2770
Db      4274 TCCAGCTCTGTATCTATGCTCTGCTGGAGAGCTTACAGAGCTGGGGCTCATACTCGG 4333
Qy      2771 HisCySerValaAnglyThrThrPheGlySerAepProgl1yCyLeuVal1LeaAngly 2790
Db      4334 CACTGACACGGCCATATGGAGCGTGACAGGACGACAGCTCTGATCAGCAATCATCAGCTGT 4393
Qy      2791 Gl1yAepProgl1yLeuProAlaAnglyLeuArgLeuGlyAenAepPheArgTyrrAsnLys 2810
Db      4394 GGGGATCTGTGACACTGCGCCAAATGGCATTCAGATTGGGACAGACTTCACCTTCACAG 4453
Qy      2811 ThrVal1ThrTyrrGlnCyVal1Progl1TyrrMetMetGlySerHisArgValSerValLeu 2830
Db      4454 ACCGTGAGCTATCTGTCAGACCCCGGCTTACTTATGAGCCCAACATCACCAACATC 4513
Qy      2831 SerCyThrTyrrAaParGlyThrThrAaAnglyThrLysProVal1CySlyAlaLeuMetCys 2850
Db      4514 CGTGACACCAAAATGGACATGAGATCAGACCGCGCCCTCTGCAAAAGCTGTCTGTGC 4573
Qy      2851 LysProProPheLeuLeuPheAnglyLysValValGlySerAepPheMetTyrrGlySer 2870
Db      4574 AACCAAGCTCTCTCCAGTCCAAATGGAAGGTGAGGGGTGACGACTTCGATGGGGTGC 4633
Qy      2871 SerVal1ThrTyrrAlaCyLeuGluGlyTyrrGlnLeuSerLeuProAlaValPheThrCys 2890
Db      4634 AGCTTAAAGCTTACAGCTGTGTGATGGCTTACCAAGCTTCCACCTGGCCATCTGTCTGC 4693
Qy      2891 Gl1yLysAnglySerTyrrThrGlyLysLeuProGlnCysPheProVal1PheCyGlyAaP 2910
Db      4694 GAAGGGGTGAGTGTGGAAGGAGAGAGTCCCTCAGTGTCTGC-GTCTTC----- 4743
Qy      2911 Progl1yVal1ProserAArgLysArgGluAaPArgLysPheSerTyrrArgSerSerVal 2930
Db      4743 ----- 4743
Qy      2931 SerPheSerCyHisLeuProProLeuValLeuValGlySerProArgArgPheCyGlnSer 2950
Db      4744 ---ATCAGTGCAGACCCCACTGCTATGATGAGATCTCTCAGAGAGAACTGCCAGGCC 4800
Qy      2951 AaPgl1yThrTyrrSerGlyThrGlnProSerCySileAaPProThrLeuThrThrCyAla 2970
Db      4801 GATGGATATGAGATGGATGATCCAGCCCACTGTATAGNNNNNNNNNNNNNNNNNNNNNN 4860
Qy      2971 AaPProGlyVal1ProGlnPheGly1LeuGlnAaAaSerGlnGlyTyrrGlnValGlySer 2990
Db      4861 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4920
Qy      2991 ThrValLeuPheArgCyGlnLysGlyTyrrLeuLeuGlnGlySerThrThrArgThrCys 3010
Db      4921 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4980
Qy      3011 LeuProAaenLeuThrTyrrSerGlyThrProProAaPArgCyVal1ProHisCyAArgLys 3030
Db      4981 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5040
Qy      3031 ProGluThrProThrHisAlaAaenValGlyValLeuAaPLeuProSerMetGlyTyrrThr 3050
Db      5041 CCAGAGACCCAGCGCATGACAGATGAGAGCCATCGATCTTCAGCTTTGGCTACACT 5100
Qy      3051 LeuLeu1-ThrProAlaArgArgAlaSerProSerArgValAlaProserThrAlaProAl 3070
Db      5101 TTAGCTTACACTGACCTGCACCCGAGATTTTCTTCTGCTGTGATCTGAACACAGAGAGCTGT 5160
Qy      3070 AaArgMetAla1aGlyGlnAlaSer 3079

```

```

Db      5161 AAAGCAGATATGAATGACAGAGAAAGT 5188
RESULT 3
AK081081
LOCUS
DEFINITION
AK081081 3882 bp mRNA linear HTC 03-APR-2004
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:B930082409 product:CTB and Sushi multiple
domains 1, full insert sequence.
ACCESSION
AK081081.1 GI:26099665
VERSION
HTC; CAP trapper.
KEYWORDS
Mus musculus (house mouse)
MUS MUSCULUS
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Aizawa, J., Nishi, K., Kikunishi, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, K., Matsuno, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3882)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotoh, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,

```

COMMENT
 Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: <http://genome.gsc.riken.jp/>
 URL: <http://fantom.gsc.riken.jp/>
 Location/Qualifiers
 1. 3882
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM DB:B930082H09"
 /db_xref="taxon:10090"
 /clone="B930082H09"
 /issue_type="cerebellum"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"
 233. 3656
 /note="CUB and Sushi multiple domains 1 (MGD|MG1:2137383,
 GB|NM_053171, evidence: BLASTN, 99%, match=3606)
 putative"
 misc_feature
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3882
 Score: 4805.50 Matches: 842
 Percent Similarity: 82.54% Conservative: 179
 Best Local Similarity: 68.07% Mismatches: 200
 Query Match: 28.29% Indels: 16
 Ds: 3 Gaps: 2
 US-10-016-248-2 (1-3104) x AK081081 (1-3882)
 QY 284 GYIIEPFOALATYTGIVARARGGLUGLYSERARGPHEHSHISGLYASPTIRLEUlys 303
 DB 2 GGCAATCCAGCCTATGGAGAGCGAGCTGGCAGCGCTTCTTGACATGGAGACAGCGCTCAC 61
 QY 304 PHEGLUCYSGINPROALAPHEGLULEUVALGILNYSALAIETHCYSGINLYASAP 323
 DB 62 TTTCAGTGGCCAGCAGCGCTTGAGCTGGTGGAGAGAGAGATTAATCTCCAGAAAAC 121
 QY 324 ASNGINTPSEALALALYSPROGLYCYVALPHESECYSPHEPHEAPHEPHER 343
 DB 122 AACCAAGTGTGGGCAACAAGCCAGCTGTGTTTCATGTTCTTCAACTTCACGGCA 181
 QY 344 PROSERGLYVALLEUSERPROANTYRPROGLYASPHYRGILYASPHISLEUHSYs 363
 DB 182 CCTCTGGATCATCCTCTCGCCAAATATCTGAGAGATGAGCAACAACATGAACCTGT 241
 QY 364 VALTPHEULILEUALATGPROGLUSERARGILEHISLEUALAPHEASNPILASAP 383
 DB 242 GTGTGTTGATTTATCTAGAGCTGGAGCGGATTCACCTCATCTTCAATGATTCGAT 301
 QY 384 VALGILUPROGLINPHEAPHELEUVALILEYASGLYALATHRALAGLUALAPROVAL 403
 DB 302 GTGGAACCTCACTTGAATCTCTCCGTCGTAAGAAGATGGAGATTTCATCATCAACAGTC 361
 QY 404 LEUGLYTHRPHESEGLYANGILNEUPROSERSEITLTHSERSEGLYHISVALA 423
 DB 362 CTGGGAACCTTCTCTGGCATGAGTGGCTGGACACACTGGCCAGCAGTGGACATAGTA 421
 QY 424 ARGLEUGLUPHEGLINHPHEASPHISSEITRTHGLYLYSARGLYPHEASNPILTHRPHE 443
 DB 422 CGCTGAGATTTCACTGATCATCTCCACACTGGCAGAGTTTCAACATCAATATATACC 481
 QY 444 THRPHEARGHISANGILUCYSPROASPPROGLYVALPROVALASGLYLYSARGPHEGLY 463
 DB 482 ACATTTGTGCAGAACAGAGTCCACGACCTTGGGATTCCTGTAAATGACGGGGTTTGA 541
 QY 464 ASPSERLEUGLLEUGLYSERSEITLESERPHELEUCYASPPGLUGLYPHELEUGLYTHR 483
 DB 542 GACAGGTTTCTGGTGGAAATTCAGTGTCTCTCACTGATGATGATGCTTTGGAACT 601
 QY 484 GINGLYSERGLINHPHEHLEUHPHEVALLEUVALLEUGLYSERVALIATRPHEASRALA 503
 DB 602 CAGGTTCTAGTCATCATCAGTGCATCTTGCAGAGAGAAAGTGTCTGGAGCTTACT 661
 QY 504 VALLEUARGCYSGILUALAPROCYSGILYHISLEUHTHSEPROSERGLYTHIRLEU 523
 DB 662 GTCCCTCGTGTGAATCTCATGTGTGTGGACCTGACCGCTTCCAGTGGGTCAATATTA 721
 QY 524 SERPROGLYTRPPOGLYPHETRYLYSAPALALEUSERCYALATRPVALILEGLUALA 543
 DB 722 CTTCCAGAGATGGCGGGATTAACAAGATTTCTTAACCTGGAATGGGTCAATTGAAGA 781
 QY 544 GINPROGLYTYRPROILEYSLIETHRPHEAPAPGPHELYTHGLIVALANTYRASP 563
 DB 782 AAMCAGACATTTCCATCAAGATTAACATTCGACAGGTTCCAGACAGAAATCAATATGAT 841
 QY 564 THIRLEUGLIVALARGASPIYARGTHYRYSERAPALPROLEULIAGLYVALTYRHISGLY 583
 DB 842 ACCCTGGAGATACGGGATGGCCACACAGCTCATCCCATCATAGAGAGATCACCGC 901
 QY 584 THIRGINVALPROGLINPHEULEULESERTHSERANTYRLEUHYRLEULEUPHESE 603
 DB 902 ACCCAGGACCCACAGATCTCTCATCGACAGCAACATCAATGATCTGCTTTAACCACT 961
 QY 604 ASPHYSEPHISSEASPILEGILYPHLEGLINLEUARGYRGLUTRILETHLEGLIN 623
 DB 962 GATPACAGCGGGCTAGAGTGGTCTCTCATCCATGAGATGAGATGATCTTGAATCC 1021
 QY 624 ASPHISCYLEUASPPROGLYIILEPROVALANGILYNARGHISGLYASNPAPHE 643
 DB 1022 GACTCTCTGTGAGACCCAGCATCCCTGTAATGATGATGACGCCCATGACATGACTTGGT 1081
 QY 644 VALGLYALALEUVALTHPHESECYASPPROSERGLYTRTHIRLEUSERASPIGLYPRO 663
 DB 1082 ATCAGGTCTACAGTACCTTCACTGACCTGAGTCCCTGGATACACTCATGATGACATGCC 1141
 QY 664 LEUGLUCYSGIUPROASNPHEGLINTPSEARGLALALEUPROSECYSGILUALALEUCYs 683
 DB 1142 CTCTCTGTGTGAGAAACCATCATGAGGAAACCAACGCTTGCACAGCTGTGATGCTGTGT 1201
 QY 684 GLYGLYPHEILEGLINLYSERSEGLYTHIRILEUSERPROGLYPIEPROASPPHE 703
 DB 1202 GAGGCTCATCATCTGAGAAAGTGGAGCTGCTTCTCCAGATTTCCAGACTTTAT 1261
 QY 704 PROASNPANLEUANCYRTHTRPILILEGLUTHSERHISGLYLYSGLYVALPHEPHE 723
 DB 1262 CCAAACTCTGAACTGATGACATGACATGAACTGCTCATGAGCAAGGAGTGCAGATG 1321
 QY 724 THRPHEHISHPHEHISLEUGLUSERGLYHISAPTYRLEULEULIETHRGILUA 743
 DB 1322 AATTTCAACCTTTTCACTTGAAAGTCCCAAGACTATTTGCTGATTCACAGAGATGG 1381
 QY 744 SERPHEHISHPHEHISLEUGLUSERGLINLEUTHRGILYSERARGLEUPROALAPRO 763
 DB 1382 AGTTTCTCAAGCAAGTACAGCAGCTCATCTGGTCACTGCTTCAACCATTAAGGCT 1441
 QY 764 GLYLEUHYRGLYANPHEHISLEUGLUSERGLYHISAPTYRLEULEULIETHRGILUA 783
 DB 1442 GGTTTGTGGAAACTTCACTGCAACATCAGGTTTATCTGATCTTCACTTCTCAT 1501
 QY 784 GLUGLYPHEASNPILTHRPHESEGLYTYRASPGLINUPROCYSGILUGLUPROGLYVAL 803
 DB 1502 GAAGGCTTCAACATTAAGTTTGCAAGATVAGCTGGAGACCTTGATGATCACTTGAGATC 1561
 QY 804 PROALATYRSEITLARGLYSGILEUGLINPHEGLYVALIAGYASPTIRLEUTHRPHE 823
 DB 1562 CTGCTCTGACCGCGAGATTTGGGTTTCACTTGGTGTGGGAGAACCTTGCTTCAACC 1621
 QY 824 CYSPHEPPOGLYTYRARGLEUGLUGLYTHRALAARGILETHRCYALEUGLYGLYARGARG 843

Db 1622 TGCTTCCAGGATACCGCTAGAGAGTGCACCAAGCTTACCTGCGGTGGGAGC 1681
 Qy ArgLeuTPSerSerProLeuProArgCyValAlaGluCyGlyYanSerValThrGly 863
 Db 1662 CGAGGTGAGAGCAGCTCTGCGCAAGGTGTGTGGCTGAATGGAGCAAGCTCAAGGA 1741
 Qy ThrGlnGlyThrLeuLeuSerProAnpPheProValaIenTYrAsnAsnAsnHieGluCyS 883
 Db 1742 AATGAGAGACATTACTCTCTCCAAATTTCCATCCATTAATGATTAACATTAAGT 1801
 Qy IleTYrSerIleGlnThrGlnProGlyYleGlnLeuValaArgAlaPheGlu 903
 Db 1802 ATCTATAAATAAGAAACAGAACCGGAGGAGATCCATCTCAGAGCTGCAACCTTCAG 1861
 Qy LeuSerGluGlyAerValLeuValaIleTYrAspGlyAanAsnAsnSerAlaArgLeu 923
 Db 1862 CTCTTCGAGGAGACCTTAAAGGTTTACATGAGAAAGACAGCTCTCAAGGCTACTG 1921
 Qy GlyValPheSerHieSerGluMerGlyValThrLeuAnSerThrSerSerLeu 943
 Db 1922 GGTGCTTACAGAGAGGGAATTCATGGGGCTGTGCTCAACAGACCTCCACACCTG 1981
 Qy TripleAspPheIleThrAspAlaGluAnThrSerIleGlyPheGluLeuHiePheSer 963
 Db 1982 AGGCTGAGTTCACACCAACGAGTCCGACACCGCCCAAGGCTTCACCTCACTACAC 2041
 Qy SerPheGluLeuIleValCyGlyLeuAspProGlyThrProValPheGlyTYrIleValHie 983
 Db 2042 AGTTTGAACCTAGTAAATGTGAGATCCAGGACATCCCTAATGATGATACAGATCCGT 2101
 Qy AspGluGlyHiePheAlaGlySerSerValSerPheSerCyAspProGlyTYrSerLeu 1003
 Db 2102 GATATGCGCACTTACAGACACTAGTCTCTACAGCTGACACCGGGCTAGCGATG 2161
 Qy ArgGlySerGluGluLeuLeuCyLeuSerGlyGluArgArgThrTrpAspArgProLeu 1023
 Db 2162 CATGGACAGATACCTGACCTGCTTGAAGTGGGACCGCAGGCTGGGACAAACCTATG 2221
 Qy ProThrCyValAlaGluCyGlyGlyThrValaArgGlyGluValaSerGlyGluValLeu 1043
 Db 2222 CTTTCCTGTGTGGCTGAATGTGTGTCTTGTCCATGACGACCATCGGGCGCATCTG 2281
 Qy SerProGlyTYrProAlaProTYrGlnHieAsnLeuAnCySileTYrThrIleGluAla 1063
 Db 2282 TCTCTGCTACCCGCCCATATACACAACTTCTCACTGACCTGGAACCATAGAGCC 2341
 Qy GluAlaGlyCyThrIleGlyLeuHiePheLeuValaPheAspThrGluValaHieAsp 1083
 Db 2342 GATCCTGGAAGACATCAGCCTCATTTTCAATGTGTGACACTGAAACGGCGCATAC 2401
 Qy ValLeuAlaGlyIleTPAspGlyProValaGluSerGlyValaLeuLeuGluLeuSerGly 1103
 Db 2402 ATCTTCAGAGTCTGGGATGGCCTGTAGACAGCAATCTGCTGAAAGATGGAGCGCT 2461
 Qy ProAlaLeuProValaAspLeuHieSerThrPheAnSerValaIleGluPheSerThr 1123
 Db 2462 TCGGCTCACTGAGAGATCAGACGACCTTCACTCACTCACTCACTCACTGATGACGT 2521
 Qy AspPhePheThrSerIleGlnGlyPheAlaIleGlnPheSerValaSerThrAlaThrSer 1143
 Db 2522 GACTTCTTCATCAGAAAGTCCGGCTTCCATCCAGTTTCCATTCATTCATTCATTCACC 2581
 Qy CyAsnAspProGlyIleProGluAnGlySerArgSerGlyAspSerTrpGluAlaGly 1163
 Db 2582 TGCATATGACCTGGGATGCTCAGAAATGTAACCCCTGCGGTGACAGCGGGAACCTGGA 2641
 Qy AspSerThrValaPheGlnCyAspProGlyTYrAlaLeuGlnGlySerAlaGluLeuSer 1183
 Db 2642 GACACCATCACCCTTCACTGATGACCTTGATGACCTGCAAGGACCGCAAGATACCG 2701
 Qy CyValaIleGlyLeuAnArgPhePheThrGlnProSerProProThrCyAlaIleAlaPro 1203
 Db 2702 TGTGTGACGTGAACACGTTTCTTGTGGAGCCAGATCTCTCTGTGATAGCTCT 2761

Qy 1204 CySGlyGlyAspLeuThrGlyProSerGlyValaIleLeuSerProAnTYrProGluPro 1223
 Db 2762 TGTGTGGAATCTGACAGGCTCTGTGAGTGAATCTATCCCAATACCAAGCA 2821
 Qy Tyr-ProProGlyIleGluCyAspTrpValaIleValaSerProAspTYrValaIleAla 1243
 Db 2822 TACCCCTCTGGAAGAGGTGCGATGGAGGATTAAGTGAAACCAAGCTTGTGATTC 2881
 Qy AlaValaPheAnIlePheAnLeuGluProGlyTYrAspPheLeuHieIleTYrAspGly 1263
 Db 2882 CTGATATCTCAAAAGTTTACATGAGAGCCAGATTAATGATCTCTGATATCTATGAAG 2941
 Qy YArgAspSerLeuSerProLeuIleGlySerPheTYrGlySerGluLeuProGlyYArgIle 1283
 Db 2942 GGAGATCTCAACAGCCACTGATTTGAAGTTTCCAGGCTTCAAGCCCAAGAGGAT 3001
 Qy eGluSerSerSerAnSerLeuPheLeuAlaPheArgSerAspAlaSerValaSerAnAl 1303
 Db 3002 CGAGAGCAGTGTAAACAGCTCTTCTGGCATTCAGAGATGATCTTGTGGCTGTCTC 3061
 Qy aGlyPheValaIleAspTYrThrGluAnProArgGluSerCyPheAspProGlySerIle 1323
 Db 3062 GGGATTCGCCATTGAATTAAGAAACCCCGGGAAGCTTGTGACCTCGGGAACAT 3121
 Qy eLYAsnGlyTYrThrArgValaGlySerAspLeuValaLeuGlySerSerValaIleThrTYrCY 1343
 Db 3122 AATGAATGGACAAAGATGGAACTGACTTAACTGGGCTTACAGTACTTCACTG 3181
 Qy SHieGlyGlyTYrGluValaGluGlyThrSerThrLeuSerCySileLeuGlyProAspGly 1363
 Db 3182 TGACTCTGTATCAAGATTTGATGATCCCTGCTCATTAAGTGTGACAGGGCTGATG 3241
 Qy YLYAspProValaIlePheAnProArgProValaCyThrAlaProCySGlyGlyIleTYrVala 1383
 Db 3242 GAAACCGTCCGAGGAGCGGACACTGCTGCTGCAAGACCTGTGGAGCGCATGTAC 3301
 Qy IGLYSerAspGlyValaIleLeuSerProAnTYrProGluAnTYrThrSerGlyGlnIle 1403
 Db 3302 GGGCTCGAGGGGAGTGTGTGTGACCAAACTACCTTCATTAACAGCTGGAGAT 3361
 Qy eCyLeuTYrPheValaIleThrValaProLYAspTYrValaIlePheGlyGlnPheAlaPhe 1423
 Db 3362 GTGATCTATTCATCAAGGCGGCCAAGAAATGTGTGATTTGACAGTTGCTATTT 3421
 Qy eHieThrAlaLeuAnAspValaIleGluValaHieAspGlyHieSerGlnHieSerArgLe 1443
 Db 3422 CCAGACTGCACTGAACGACTTGAGAGATTTGATGAAACACACCCCGCAGCGCT 3481
 Qy uLeuSerSerLeuSerGlySerHieThrGlyGluSerLeuProLeuAlaThrSerAnGly 1463
 Db 3482 TCTGAGTCTCTCTGTGTCTCCATTCAGGTGAAGAACTGCACTGTGTATCATCA 3541
 Qy nValLeuIleLYAspPheSerAlaLYSlyLeuAlaProAlaArgIlePheHiePheValaTY 1483
 Db 3542 GATTCGTCTGATTCACGCGCAAAAGCGAGCTTCTGACGGGTTTCCACTTCGTCTA 3601
 Qy rGlnCySerSerValaProGluProArgTYrGlyLYArgLeuGlySer 1508
 Db 3602 CCAGAGAGAGAAAGTCCCGGCGCTGACTTGTGAGACAATGTGAGGTGAGCTAC 3661
 Qy 1492 rGlnCySerSerValaProGluProArgTYrGlyLYArgLeuGlySer 1508
 Db 3662 ATATTGT-----GTAACAAAGAGAGGTGCTCCCGGCTCC 3698

RESULT 4
 LOCUS AY407074 5824 bp DNA linear GSS 15-DEC-2003
 DEFINITION Pan troglodytes CSMD1 gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY407074
 VERSION AY407074.1 GI:39763045
 KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 REFERENCE Mammalia: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 AUTHORS 1 (bases 1 to 5824)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B., Fierliera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J., Adams, M.D. and Cargill, M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 5824)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civeille, D.R., Lu, F., Murphy, B., Fierliera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J., Adams, M.D. and Cargill, M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment
 FEATURES
 source location/Qualifiers
 1..5824
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 <1..>5824
 /gene="CSMD1"
 /locus_tag="HCM2769"
 gene
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 5824
 Score: 4001.00 Matches: 720
 Percent Similarity: 63.68% Conservative: 176
 Best Local Similarity: 51.17% Mismatches: 453
 Query Match: 23.56% Indels: 58
 DB: 9 Gaps: 1
 US-10-016-248-2 (1-3104) x AY407074 (1-5824)
 QY 1320 ProGlySerIleuYsaemGlyThraYvaIGISeRspLeuYsleuGlySerServal 1339
 DB 2 CCAGGAATATTAATGAATGGAGCAAGATGGAACTTCAAGCTTGCTCCACATC 61
 QY 1340 ThrTYTYTCySHISGLYGLYTYRGLUValIGLUGLYThSerThreusSerCysIleu 1359
 DB 62 ACCTACCAAGTGTACTCTGGCTACAAAGATCTTGACCCCTCATCCATCACCTGTGATT 121
 QY 1360 GlyProAspGlyLySPProValTTPAsnAsnProArgProValCysThraIaProCysGly 1379
 DB 122 GGGGCCCATGGGAAACCTCTCTGGGACCAAGTCTGCTCCGCAATGCTCCCTGTGA 181
 QY 1380 GlyGlnTYValIGlySerAspGlyValValIleuSerProAsnTYProGlnAsnTYThr 1399
 DB 182 GGCACAGTACACGGGATACAAAGGGTATTTATTCACCAAACTACCCCATATATACACA 241
 QY 1400 SerGlyGlnIleCysLeuTYRPhValThValProLYAspTYValValPheGlyGln 1419
 DB 242 GCTGCTCAATATATGCTCTATCCATCAACGGTACCAAGGAATTCGNNNNNNNNNNNN 301
 QY 1420 PheAlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGln 1439
 DB 302 NNN 361
 QY 1440 HisSerArgLeuLeuSerSerIleuSerGlySerHisThraGlyGlnSerIleuVal 1459
 DB 362 CAGGCCAGACTTCTACGCTCGCTCTGGGCTCTCACTCAAGGGGAAATGCTTGGCT 421
 QY 1460 ThSerAsnGlnValIleuIleuYpHeSerAlaYsGlyLeuAlaProAlaArgGlyPhe 1479

DB 422 AGTCAATCAAAATTCGTCTCCGATTCAGTCAAAAGCGGTCTCTGCCCCGGGCTTC 481
 QY 1480 HisPheValTYRGLNAlaValProArgThrSerAlaThrGlnCysSerServalProGlu 1499
 DB 482 CACTTCGTGATTAAGCTGTTCTCTGACAGGACCAACCAAGCAAGCTGTCGCCGAG 541
 QY 1500 ProArgTYRGLYLYsaRGLeuGlySerAspPheSerValGlyValIleValArgPheGlu 1519
 DB 542 CCCAGATACCGAAGAGAAATGGTCTGAGTTTGGCCGGGCTCATCTCCATTCCAG 601
 QY 1520 CysAsnSerGlyTYRValIleuGlnGlySerProGlnIleGluCysLeuProValProGly 1539
 DB 602 TGCACCCGGGATACCTGCTTCAAGGCTTCACGGGCTCCACTGCCAGTCCCGCCAT 661
 QY 1540 AlaLeuAlaGlnIlePheAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeu 1559
 DB 662 GCCTTGACACAGTGAACACACAGATCCCGACTGTGTGTCCTTCAGTGGCAATTC 721
 QY 1560 ThrGluArgArgGlyThrIleLeuSerProGlyPheProGluProTYLeuAsnSerLeu 1579
 DB 722 ACTCAAGAAAGAGTACAATCTGTCCTCCCGCTACCTGAGCCATACGAAACAATCTG 781
 QY 1580 AsnCYValATrPheIleValValProGluGlyAlaGlyIleGlnIleGlnValValSer 1599
 DB 782 AACTGATATGAAGATCTACTTACGAGGAGCTGGGANNTTCAATCCAGTCAAGTACAGT 841
 QY 1600 PheValThrGluGlnAsnTrpAspSerLeuGluValPheAspGlyAlaAspAsnThraVal 1619
 DB 842 TTTGCCACGAGCAAGATCGGAGCTCCCTTGATCCAGATGGTGGGATGTGACCCCA 901
 QY 1620 ThrMetLeuGlySerPheSerGlyThrThraValProAlaLeuAsnSerThSerAsn 1639
 DB 902 CCGACACTGGAGAGCTTCTCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 961
 QY 1640 GlnLeuTYLeuHisPheTYSerAspIleSerValSerAlaIleGluPheIleuGlu 1659
 DB 962 NNN 1021
 QY 1660 TYRlySThrValIGlyLeuSerSerCySPProGluProAlaValProSerAsnGlyValLYs 1679
 DB 1022 NNNNNNNCTGTAGGCTTGTGTCGATGCAAGAACACACCCCTCCACCAACAGATCAAA 1081
 QY 1680 ThrGlyGluArgTYRLeuValAsnAspValValSerPheGlnCysGluProGlyTYRVal 1699
 DB 1082 ATCGAGATCGGTACATGATGTAACAGATGCTCTCTCCATGAGCGCCGGGTACACC 1141
 QY 1700 LeuGlnGlyHisAlaHisIleSerCysMetProGlyThraValArgATrPAenTYPro 1719
 DB 1142 CTGACAGGCGCGTCCACATTTCTGTATGCAAGGAGACCGTGGCGGTGGAATATCCG 1201
 QY 1720 ProProLeuCYsIleAlaGlnCysGlyGlyThraValGluGluMetGluGlyValIleLeu 1739
 DB 1202 TCTCCCTGTGATTCGAACTGAGAGGAGACCTGAGCACTTGGGTGTGTGATCTCTG 1261
 QY 1740 SerProGlyPheProGlyAsnTYRProSerAsnMetAspCysSerTrpLYsIleAlaLeu 1759
 DB 1262 AGCCCCGGCTTCCGGGTTCTTACCCCAACATTTAGCTGACACCTGAGAGATCTCTTA 1321
 QY 1760 ProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrgluProAsnHisAsp 1779
 DB 1322 CCCATCGGCTATAGTGCACATATTCAGTTTCTGAATTTTCTACCAAGCTATATCATGAC 1381
 QY 1780 TYRleGluIleArgAsnGlyProTYRGLUThSerArgMetMetGlyArgPheSerGly 1799
 DB 1382 TTCTTAAATTCAAAATGACCTTACACACAGGCCCAATGATGACAAATTTTACCGGC 1441
 QY 1800 SerGluLeuProSerSerIleuLeuSerThSerHisGluThThraValTYRPhHisSer 1819
 DB 1442 ACGATATCCCGCGGCGCTGCTACACACAGCATGAACCTCATCACTTATATAT 1501
 QY 1820 AspHisSerGlnAsnArgProGlyPheLYLeuGluTYRGLNAlaTYRGLUleuGlnGlu 1839
 DB 1502 GACCAATTCGAAACCGGCAAGATTAACTTGTTCACCAAGCTTATGATTAACAGAAC 1561

1840 CybProAspProGluProPheAlaAsnGlyIleValaArgGlyAlaGlyTyrAsnValGly 1859
1562 TGTCAAGTTCACCCCATTTTCAGAAATGGGTACATGATCAATCACTGATTAAGGGTGGG 1621
1860 GlnSerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThr 1879
1622 CAATCAATATCTTCCAGGTGTATCTGGGTACATTCGAATAGGCCATCTGCTCTACT 1681
1880 CysGlnHisGlyThrAsnArgAsnTyrPheHisProLeuProGlyCysGluValProCys 1899
1682 TGTGACACGGGATCAACAGAACTGGAATCACTTTCCAGATGATGAAGCCCTTGT 1741
1900 GlyGlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyr 1919
1742 GGGTACAACTGATCTTCTCAGAACGGCACATCACTCTCCCTGGCTTCTCGATGAGTAT 1801
1920 SerSerSerGlnAspCysValTyrPheIleThrValProIleGlyHisGlyValaArgLeu 1939
1802 CCGATCCGTAAGAGCTGATTTGGCTCATCACTAGCTCCGGGGCACGGAGTTTACATC 1861
1940 AsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyPro 1959
1862 AACTTCACTCTGTACAGACGGAGCTGTCAACGATTAATGCTGTGTTGNNNGTCCC 1921
1960 GlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaValGlyThrValGln 1979
1922 GATCAGAACTCAACCCAGCTGGAGTTTTCAGTGCAACACAGCTCTGANNNNNNNTAT 1961
1980 SerSerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAlaThrGlyIlePhe 1999
1982 AGCTCCACCAACCAAGTCTGCTCAAGTTCACAGCCACTTTCAATGAGAGGTTCTTT 2041
2000 AlaIleAlaPheSerAlaTyrProLeuThrLysCysProProThrIleLeuProAsn 2019
2042 GTCTCTCAATTTCCAGNN 2101
2020 AlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIleValaArgTyrArgCys 2039
2102 NNN 2161
2040 LeuProGlyPheThrLeuValaGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeu 2059
2162 CACCCCGGGTACACTTGGTGGGACCGACACTNNAGCTTCCAGCTCAGTTCACAGTTG 2221
2060 GlnPheGluGlyProProPheIleCysGluValHisCysProThrAsnGluLeuLeuThr 2079
2222 CAGTTGAGGGGTTCCCTCCCAACATGTGAAGNNNNNNNNNNNNNNNNNNNNNNNN 2281
2080 AspSerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThr 2099
2282 NNN 2341
2100 CysSerThrLeuValaArgValGluProAspTyrAsnIleSerLeuThrValaGlyTyrPhe 2119
2342 NNN 2401
2120 LeuSerGlyLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerPro 2139
2402 NNN 2461
2140 LeuLeuValaIleLeuSerGlyAsnTyrSerAlaProLeuIleValIleThrSerSerAsn 2159
2462 CTGTGTGTAATCTTAAGTGGAAATCATGAACTCAATCAATTTTACAGAGGAGTAT 2521
2160 SerValTyrLeuArgTyrPheSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArg 2179
2522 CAGTTAATCTCCGCTGCTGCTCAGTACATGCCACAGTAAGAAAGATTTCAATTCGC 2581
2180 TyrSerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyIleLeuGlyGln 2199
2582 TATCAGCACACTTACTGACAGTTGACCCACCGCTGAAGAAATGGGGATATTCTAAACAGA 2641

2200 ThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuVal 2219
2642 ACTCAGAGACCGGTGGAGAACCAAGTCAATATTTTTCAGAACCTGATACCGAATGTC 2701
2220 GlyHisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrPheSerGluAla 2239
2702 GGCCACAGCAATGCACTGTAGACGAACCACTTGCGATGTACAGGTGGACTCCCTC 2761
2240 IleProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetVal 2259
2762 ATGCACCTTGGCCAGGCTGTCTCTGTGAATCCCAAAATCCCGAGAAAGGTTCAATT 2821
2260 PheGlyLysGlyTyrThrValaGlyThrLysAlaValTyrSerCysSerGluGlyTyrHis 2279
2822 ACCGGAAACAGATTCATTGGACAGTAAGTGTCTATGATATCTATGAGGCTTCAAG 2881
2280 LeuGlnAlaGlyAlaGluAlaThrAlaGlyCysLeuAspThrGlyLeuTyrPheSerAsnArg 2299
2882 CTGGAATCCAGCCAGCAAGCAACAGCCGTGTGTCAAGAAAGATGGGCTGTGAGTAAAG 2941
2300 AsnValProProGlnCysValProValThrCysProAspValSerIleSerValGlu 2319
2942 GGGAAAGCCGCCACGTGTAAAGCCGTCTTGGCCCAAGTGAAGCTCAAGCTTCAGAA 3001
2320 HisGlyArgTyrPheGlyLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeu 2339
3002 CATGTCACTGGAGAGCTGTTCAGATCTCTGAAATAGTACGGTGTCAAGTATGCTG 3061
2340 IleCysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGly 2359
3062 AGTGCAGCTCTGTGTACTTCTTAGAGGGCCGAGGCTCTCGACAGTGTCAAGGCCAATGG 3121
2360 LysThrSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuPro 2379
3122 ACGTGAACGTAGAGATGAGAGGCAAGCTGTGCA----- 3157
2380 IleProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePhe 2399
3157 ----- 3157
2400 SerCysAsnSerGlyTyrThrLeuValaGlySerArgValaArgGlyCysMetAlaAsnGly 2419
3157 ----- 3157
2420 LeuTyrPheSerGlySerGluValaArgCysLeuAlaGlyHisCysGlyTyrProGluProIle 2439
3158 -----GCTGGCCACTGCGGTTCCCAAGCCCAATT 3187
2440 ValaAsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValaIleTyrGlnCys 2459
3188 GTGAACGATCACTTATGTGAGATGAGCTTCACTGATACAGACAGGAGTGTATACCAAGTGC 3247
2460 AsnAlaGlyPheArgLeuIleGlyMetSerValaArgIleCysGlnGlnAspHisIleTyr 2479
3248 AATCTGTGTTCCGGCTGTGGGAATCCGTGAGATAGCTTAACAAGACCAAGATGG 3307
2480 SerGlyLysThrProPheCysValProIleThrCysGlyHisProGlyAsnProValAsn 2499
3308 TCTGGACAAACGCTGTGTGTGTCCCATCAATGTGTGTCAACCTTGAAACCCCGCCAC 3367
2500 GlyLeuThrGlnGlyAsnGlnPheAsnLeuAspValaValLysPheValaCysAsnPro 2519
3368 GGAATTCATTAATGAGATGAGTTCATCACTGAATGACATTCGTAATTCACCTGACACAG 3427
2520 GlyTyrMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTyrPheSerAsp 2539
3428 GGTATATTGTCGACGGCGTGTCTCGAGCCAGTGTGAGCAACGGCCAGTGAAGTACG 3487
2540 MetLeuProThrCysValGlyIleIleAsnCysThrAspProGlyHisGlnGluAsnSerVal 2559
3488 CTTCTGCCACAGTGTCAAGTGTGAATCTTCTTATCTCAGGCTTTTNTGAAATGNNNNN 3547
2560 ArgGlnValHisIleAsnSerGlyProHisArgPheSerPheGlyThrThrValaSerTyrArg 2579

Alignment Scores:

Pred. No.:	1244-95	Length:	744
Score:	1284.00	Matches:	233
Percent Similarity:	96.79%	Conservative:	11
Best Local Similarity:	92.37%	Mismatches:	8
Query Match:	7.56%	Indels:	1
DB:	7	Gaps:	0

US-10-016-248-2 (1-3104) X CF744273 (1-748)

277...MSDEGGLIADGIYALDGLUALAINTLAGLUCYSLEUASPThrgLYLeUTrpSerAsn 2298

Db
1 CATCTCCAGCGGGTGC CGAGGCCACCGCAGAGTGTCTGGATACTGGCCTGTGAGCAAC 60

[illegible]

2318

61 AGCAATGTCCTCAGTGTCTCCCTGTGACTTGTCTCGATATCAGCAGCATCAGTGTG 120

QY 2319 GIuHISG1YArqTRpArqLeuIlePhEG1uThrg1nTyvrG1nPhEG1nA1ac1nIcWcE 2332

121

221 CAGGACCCAGTATCAGTTCAGGCCAGCTGATG 180

2339 LeuIleCysAspProGlyTyrTyrTyrThrGlyGlnargValIleargCysGlnAlaAsn 2358

db
181 CTCATCTGTGACCACTACTAACCCCAAGCCTTTTTC

[illegible]

2359 GlySerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeu 2378

D**b** 241 GGGAGTGGAGCCCTGGGAGAGTCGATGCCCACTGCCAATCATCTCTGTGGAGAGCTTC 300

0v 3379 BcCTJcFmFmT-... ..

[illegible]

301 CCCACACCCCAAGTGGGCACCGCATTTGAACTATGTCTCTATGGGGCGACTGCCATC 360

OY
2399 pHeSerCvSAsnSerG]vTvtThrt.aVtJcI.vSe-t-vvJ]v-
.....

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

361 TTTCTCCGCAATTCGGATATACACTGGTGGCTCCAGGTGCGGAGTGCATGCCCAAT 420

2419 GILeUTrpSerGluValArCvSLeuAlagIvHIaCvsgGlyTbhrPrroGluPrs 3430

421

...GACACGCGGACTCCAGAGCCC 480

2439 ILevalAsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValValTyrGln 2458

[illegible][illegible]

2459 CYSASNLAGLYPHEARGLEUILEGLYMETSERVALARGILECYSGINGLNASPHISHS 2478

541 TGCAGGGCTGGCTTCCGGCTGATCGGCATGTCCTGCGCATCTGCACAGGATCATAC 600

2478 PUNGGULUT... ..

...Proteinase 2498

601 TGGTCGGCAGAC-CCTTCTGTGTGCCAATTACCTGTGACACCCAGCAATCCAGTC 659

2499 Aeng[vIeUThrg] nG[vAeeng] rPhoAnf auAeAe-u-1-w-1 x... ..

Db	660	AACGGCCCTTAAGTACGAGGAGCAGTTCACCTGACGAGCGTGAAGTTCATTGTCANCC	712
Oy	2519	Proglutinyrmechlagluclyalala	2527
Db	720	CNAGGATACATAGCTGAGGAGGCTCG	746
RESULT 6			
LOCUS	AK046385	2841 bp	mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus adult male corpora quadrigemina cDNA, RIKEN full-length enriched library, clone 8230378117 product:hypothetical Sushi domain / SCR repeat / CDP module/ CUB domain containing protein, full insert sequence.		
ACCESSION	AK046385		
VERSION	AK046385.1	GI:26091501	
KEYWORDS	HTC; CAP trapper.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Mech. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20495374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Teshiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazada, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasliwal, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
MEDLINE	5		
PUBMED			
REFERENCE	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	6 (bases 1 to 2841)		
PUBMED			
REFERENCE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hizomoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kaya, H., Kihara, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyachi, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohtsuo, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shitaki, R., Sugabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of		

DB	QY	Score	Percent Similarity	Best Local Similarity	Query Match:	DB:
US-10-016-248-2	(1-3104) x AK046385 (1-2841)					
QY	1	8.89e-90	Length: 2841			
	1229.50	Matches: 243				
	77.29%	Conservative: 19				
	71.68%	Mismatches: 44				
	7.24%	Indels: 34				
	3	Gaps: 4				
QY	21	AlaSerAsnGlnArgHisSerValGlyValGly-----ProSerGluIleu	35			
Db	1009	GGCTCCACTTCACATCAGATGTGTAACCAACCGGACGAGCGCTTCAGCCGACCAACCA	1067			
QY	36	ValIleValSerGlnIleGluLeuIleValSerArgGlyValIleValLeuMetProSerIleValAspAsn	55			
Db	1068	GTCAGAGACAGATGTGAGCTGCAAGTCTCCAGAGCCTGAAGCTGAAGCCCAAGAGACAT	1127			
QY	56	SerGlnIleValSerValIleuThrGlnValGlyValSerGlnIleValIleValSerMetCysPro	75			
Db	1128	AACAGAAAGACATGTGTGTAACCTCAGAGTGTGTGTCCCAAGAGACATATATGTGTCA	1187			
QY	76	AspProGlyIleLeuProGluValArgGlyValArgLeuGlySerAspPheArgLeuGlySerSer	95			
Db	1188	GACCTGTGGCATCTGTAAGAAAGGGAAGAAAGCTAGGCTCAACATTCAGATTAGATTCAGC	1247			
QY	96	ValGlnPheThrCysAsnGlnIleGlyValArgLeuGlnIleValSerIleValIleThrCysMet	115			
Db	1248	GTCAGATTCACTGTGCAACGAGGCTATGACCTGCAAGAGGTCCAAAGAAATCACTGTGATG	1307			
QY	116	IleValSerAspMetPheAlaValIlePheSerAspHisIleArgProValCysArgAlaArgMet	135			
Db	1308	AAAGTGGACCAATGTGTTCAGCATGGAGTGAACCAACAGCGGAGTCTCCGAGCCCGCATG	1367			
QY	136	CysAspAlaHisIleValArgGlyProSerGlyIleIleThrSerProAsnPheProIleGln	155			
Db	1368	TGTGATGCCACCTCCGAGGTCTCCGTGGATATATACATCGTCCCAACATTCCTCCATTCAG	1427			

```

QY 156 TyraaPaanaaanaaiahiScysValTrrpIleIeThrAlaIeuasProSerIysValIle 175
Db 1428 TATGACAAACATGACACCTGCTGATATATATCTGACCTCAACCTTCCAAAGGATATC 1487
QY 176 TyLeuAlaIhegluIuPheapLeuGluAargIlyTyraSPThrIeuThrValIgluAsp 195
Db 1488 AAATCGGCTTTTAAAGATTGATCTGGAGAGGGGCTATGACACCTGACACTCCGGGAGC 1547
QY 196 GtYglYgluAspGluAspGluIysThrValIeuTyMetSerGluAsnAlaCysSerasp 215
Db 1548 GGTGGGACAGATGGGACCAAGAGACATCTCTTACATC----- 1586
QY 216 SerProhiTherProGlySerAargIleProGluSerMetSerGlyAspIleThrAargIn 235
Db 1587 -----CTGACAGGTACTCTCCGTCACAGC---CTCATGCTGACACACCATCA 1634
QY 236 LysTrpThrValIeuGluIleCysAargAspIleSerSerSeraspAlaAargSerGlySer 255
Db 1635 ATGTGGCTCTCTTCAG----- 1652
QY 256 ValAArgLysSerProLysTherSerAsnAlaValIgluValAlaProGlyThrGluIle 275
Db 1653 -----AGTACACAGACAGTGCAGCTCCCTGGGATTCAGGCTTCTTACGAAGAATC 1703
QY 276 GluGluGlySerCysGlyAspProGlyIleProAlaTyrgIyAArgAargGluGlySerArg 295
Db 1704 GAACAGGGCAGTGTGGAGACCTTGACCTGCTGATGAGCCAGAGAAAGCTCTCGC 1763
QY 296 PheIshIshIgluAspThrIeuLysPheGluCysGlnProAlaPheGluIeuValIgluIn 315
Db 1764 TTTCGCATGTGACACACCTGATTAATGCCAGCTGCTGCTTGTAGCTGGGGGAG 1823
QY 316 LysAlaIleThrCysGluLysAsnAsnGlnTrpSerAlaLysLysProGlyCysVal 334
Db 1824 AAATCAATCACTGCAAAAGAAATACAGATGCTGAGCTAAGAACCAAGCTGTGTG 1880

RESULT 7
LOCUS CFS31622 666 bp mRNA linear EST 12-SEP-2003
DEFINITION UT-M-FY0-cgg-g-06-0-UT.r1 NIH BMAP_FY0 Mus musculus cDNA clone
IMAGE:30356213 5', mRNA sequence.
ACCESSION CFS31622
VERSION CFS31622.1 GI:34583590
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgs.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strauberg, Ph.D.
Email: c9apds-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: PYX-5.
Location/Qualifiers
1. 666
/organism="Mus musculus"
/mol_type="mRNA"
/scrain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30356213"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"

```

```

/lab_host="DH10B (T1 phage resistant)"
/clone_11b="NIH BMAP_FY0"
/notes="Organ: Brain; Vector: pyx-Asc; Site_1: Ecor I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGGGAACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

```

ALIGNMENT SCORES:

```

Pred. No.: 5,56e-87 Length: 666
Score: 1182.00 Matches: 210
Percent Similarity: 97.74% Conservative: 6
Best Local Similarity: 95.02% Mismatches: 5
Query Match: 6.96% Indels: 0
DB: 7 Gaps: 0

```

US-10-016-248-2 (1-3104) x CFS31622 (1-666)

```

QY 2794 GtYleProAlaAsnGlyLeuAargLeuGluYAsnAspPheArgTyraSPThrValIthr 2813
Db 3 GGGATTCAGACCAAGGATCCGGCTGGGCAATGCTTCAAGTCAACCAACACCTGTGACA 62
QY 2814 TyrgInCysValProGlyTyMetMetGluSerHisAargValSerValIleuSerCysThr 2833
Db 63 TACAGTGTGTCCCGGCTATGTATGATGAATACACAGAGTGTGTGTGCTGAGCTGAC 122
QY 2834 LysAAspAargThrTPaAsnGlyThrLysProValCysLysAlaIeuMetCysLysProPro 2853
Db 123 AAGACCGGACCTGGAAGGATACCAAGCCGGTTGGCAAACTTATCAATGATCAAACTCTCT 182
QY 2854 ProLeuIleProAsnGlyLysValIglYSerAAspPheMetTrpGlySerSerValThr 2873
Db 183 CAGCTCATCCCGAATGGGAAGTGTGGGTCTGACTTCAAGTGGGCTTAAAGGTAGC 242
QY 2874 TyraIaCysLeuGluGlyTyrgInLeuSerLeuProAlaValPheThrCysGluIyAsn 2893
Db 243 TATGCTGCTGAGAGGATACAGCTCTCCCTGCTGCTGCTCACTGCGAGGAAAC 302
QY 2894 GlySerTrpThrGlyGluLeuProGluCysPheProValPheCysGlyAspProGlyVal 2913
Db 303 GATCTCGGACCGGAGAGCTGCTCATGCTTCCCGTGTCTGCGAGACCCAGGGGCTC 362
QY 2914 ProSerAargIyAargGluAAspAargIyPheSerTyraSPSerValSerPheSer 2933
Db 363 CCCCTCGAGAGAGAGAGAGACCGAGGTTCTCTTACAGGTCATCTGTCTCTTC 422
QY 2934 CysHiSPProProLeuValIeuValIglYSerProAargAargPheCysGlnSerAspIyThr 2953
Db 423 TGCACAGCTCTTGTGTGCTGTGGTGGCTCCACGAGGATTTTGGACAGTCAACCGGAG 482
QY 2954 TrpSerGlyThrGluProSerCysIleAspProThrLeuThrThCysAlaAspProGly 2973
Db 483 TGGAGTGGACACAGCCAGCTGATAGATCCACGCTGACCAACATGTGCAATCTGGC 542
QY 2974 ValProGluPheGlyTlIeGluAsnAsnSerGlnGlyTyrgInValIglYSerThrValIeu 2993
Db 543 ATGCGCAGATTGGAATACAGAAAGCTCCCAAGGCTTACAGTGTGGAGACCGGTGCTA 602
QY 2994 PheArgCysGluLysGlyTyraSPLeuGluGlySerTrpThrAargThrCysLeuProAsn 3013
Db 603 TTCGGTGTCAAAAGGTTACTCTGCTTCAAGGATCACCACAGGACCTGCTTCCCAAC 662

```

QY 3014 Leu 3014
 DB 663.CTG 665
 RESULT 8
 LOCUS B0832212
 DEFINITION B0832212-subtracted library Mus musculus cDNA 5' similar to
 polydcm, mRNA sequence.
 ACCESSION B0832212
 VERSION B0832212.1 GI:22864267
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2809)
 Hackney,J.A., Charbord,P., Brunk,B.P., Stoeckert,C.J.,
 Lemischka,I.R. and Moore,K.A.
 A molecular profile of a hematopoietic stem cell niche
 Proc. Natl. Acad. Sci. U.S.A. 99 (20), 13061-13066 (2002)
 2224/628
 JOURNAL MEDLINE
 PUBMED
 COMMENT
 TITLE
 Department of Molecular Biology
 Princeton University
 217 Lewis Thomas Laboratory, Washington Road, Princeton, NJ 08544,
 USA
 Tel: 609 258 0605
 Fax: 609 258 2759
 Email: kmoores@molbio.princeton.edu
 These ESTs are derived from a subtracted cDNA library enriched for
 gene products expressed by a hematopoietic stem cell-supporting
 stromal cell line, AFT024.
 Seq primer: M13Reverse or T7.
 Location/Qualifiers
 1..2809
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /tissue_type="Fetal Liver"
 /cell_type="Stromal cell"
 /cell_line="AFT024"
 /dev_stage="Embryonic day 14-14.5"
 /lab_host="DH10B"
 /clone_id="AFT024-subtracted library"
 /note="Organ: Fetal Liver; Vector: Sport 1; Site 1: Sal I;
 Site 2: Not I; Two directionally cloned cDNA libraries
 were made from a hematopoietic stem cell-supporting
 stromal cell line (AFT024) and from a non-supporting
 stromal cell line (2018). Subtractive hybridization was
 performed by hybridization of the target, AFT024, single
 stranded cDNA library in psport1 to biotinylated RNA
 transcribed from the driver, 2018 cDNA library in psport2
 with inserts cloned in the complementary orientation. The
 AFT024-subtracted library contains 4,2x105 clones and is
 depleted of common housekeeping gene products eg.
 beta-actin and enriched for transcripts specific to
 AFT024. For detailed protocols and additional information
 please see our website at
 http://stromalcell.princeton.edu."

ORIGIN
 Alignment Scores:
 Pred. No.: 2,976-81 Length: 2809
 Score: 1126.50 Matches: 266
 Percent Similarity: 45.25% Conservative: 144
 Best Local Similarity: 29.36% Mismatches: 363
 Query Match: 6.63% Indels: 133
 DB: 5 Gaps: 30

US-10-016-248-2 (1-3104) x B0832212 (1-2809)

QY 2167 SerAspHisAla-----TyrAsnArgLysGly---PheLysIleArgTyrSerAlaPro 2163
 DB 56 AGTAGACCGCTCTGCTGCTGACACCAACGAGATCTCATGATTTCTGTAACCCACCA 115
 QY 2184 Tyr-----CysSerLeuProArgAlaPro 2191
 DB 116 TACACGGGAGATGGGAAAACTGTCGAAACCTGTAAATGTATAGGCTCCGAAATATCA 175
 QY 2192 LeuHisGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPheGly 2211
 DB 176 GAAATGCGCCACTCTTCTGTGAGATTATACAC---GTGGGTACTGCAAGTCACATTTTCC 232
 QY 2212 CysAsnAlaGlyTyrArgLeuValGlyHisSerMetAlaIleCysThrArgHisProGln 2231
 DB 233 TGTACAGAGGACGACGAGCTGTGTGGAGTGAAGCAACATCACTGTGTGAGACTGGCGAG 292
 QY 2232 G1YTyHisLeuTyrSerGluAlaIleProLeuGlyGlnAlaLeuSerGlyLeuPro 2251
 DB 293 -----TGGATCGCTCAGGCCGCTGCTGAAGCAATTTCTGTGCTGCCA 340
 QY 2252 GluAlaProLysAsnGlyMetValPheGlyLysGlyTyrThrValGlyThrLysAlaVal 2271
 DB 341 CCTGTCTCTGAAATGTGTGTGTGACGCGGTCCGATTCATACATACGACATAGAGTGCTG 400
 QY 2272 TyrSerGlySerGlyGlyTyrHisLeuGlnAlaGlyAlaThrAlaGluCysLeu 2291
 DB 401 TACAGGTGTATGAAGATATACCTTTG---TCTGGGATGAAGATGACG---TGCTT 454
 QY 2292 AspThrGlyLeuTyrPsrAsnArgAsnValProProGlnCysValProValThrCysPro 2311
 DB 455 GCTAGTGCTTCTGAGTCACT 508
 QY 2312 AspValSerSerIleSerValGlnHisGlyArgTyrGlnIlePheGlyThrGlnTyr 2331
 DB 509 CAGCCTGAGCACAATAAT-----AACGGCAATAATCACTTAAGTGGGCTCACCTACTT 562
 QY 2332 GlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGlyGlnArg 2351
 DB 563 TCTATTCATCTG-----TACTCTGTGAGAACGATACCTTACAGGGGCCATCC 613
 QY 2352 ValIleArgCysGlnAlaAsnGlyLysTyrPsrLeuGlyAspSerThrProThrCysArg 2371
 DB 614 CTCTTGAATGCACAGCTTCCGACGCTGGAC-----AGAGGCCACCTAGTGTCATA 667
 QY 2372 IleIleSerGlyGlyGlnLeuProIleProProIleProIleArgIleGlyThrLeuSer 2391
 DB 668 CTGTCTCTGCGGAGAGCTCCCAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
 QY 2392 ValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrLeuValGlySerArg 2411
 DB 728 ACTTTTGGGAACACAGTGTCTTACACATGCACAAAGAGGCTTACCTTGTGCTGGCTGAC 787
 QY 2412 ValArgGlyCysMetAlaAsnGlyLeuTyrPsrGlySerGlyValArgCysLeuAlaGly 2431
 DB 788 ACCATCTGATGCCAGGCCCAACGCAATGAATGAATCAAGTAACCAACAGTGTGCTGCTG 847
 QY 2432 HisCysGlyThrProGluProIle-----ValAsnGlyHisIleAsnGlyGluAsn 2439
 DB 848 TCTGTGACGAGCCGCCCAATGTGACACGCTCTCCAGAGACTGCTCACAGGCTCTTT 907
 QY 2440 -----ValAsnGlyHisIleAsnGlyGluAsn----- 2448
 DB 908 GGAGACACCGGTTTACTACTGTGCGAGATGTATACGCTGTGATTAATTTCCAGCTC 967
 QY 2448 ----- 2448
 DB 968 ATCTGCAATGCCAGGGAACTGGTTCCTCCCGCGGGCAGAGCTGTGCGCGTGCATA 1027
 QY 2449 -----TyrSerTyr----- 2451
 DB 1028 GCTCACTTCTGTGAAAAACCCCATCTGTCTTCTACAGCATCTTGAATCTGTGAGCAA 1087

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	2126	ATCCTCAACGGATATCATCCACGTACAGAGGCTTGGCTTATCTCAGACACCAACGCTATCAC	21					
QY	2816	CysValProGlyTyrMetMetGlnSerHisArgValSerValLeuSerCysThrLysAsp	28					
Db	2186	TGCAGGCCAGCGCTTGAATTAGTATGTGGGCAATGCTACACCTTC-----TGTGGGGAAAT	22					
QY	2836	ArgThrTrpAsnGlyTyrLysProValCysValalaLeuMetCysLysProProLeu	28					
Db	2240	GCCAGTGCTCTCGAGAGAAACCAATGTGCMAAACCCATTGAAATGCCAGAGCCCAAGAG	22					
QY	2856	IleProAsnGlyLysValalaIglySerAspPheMetTrpGlySerSerValThrTyrAla	28					
Db	2300	ATTTAAATGGCCCAATCTCTCTCCGTAGCTTTAGATGAGACAAACCATCATCTACTTT	23					
QY	2876	CysLeuGluGlyTyrGlnLeuSerLeuProAlaValPheThrCysGluGlyAsnGlySer	28					
Db	2360	TGTATACCGGGCTTCGCGCTCGAAGGCCCAATCCCTGACCTGTTAGAGACAGGTGAC	24					
QY	2896	TrpThrGlyGlnLeuProGlnCysPheProValPheGlyGlyAspProGlyValProSer	29					
Db	2420	TGGAGATATNGATCCCTCTCTTGTGATGCCATCATCGACGTACACCAAGCCCATTTAA	24					
QY	2916	ArgGlyArgArgGlnLeuAspArgGlyPheSerTyrArgSerSerValSerPheSerCysHis	29					
Db	2480	AATGTTTTCGTAAAGGTGCGGATATCAGATGACGTACGTCATGATCATATAGCTCTTC	25					
QY	2936	ProProLeuValLeuValGlySerProArgAspPheCysGlnSerAspGlyThrTrpSer	29					
Db	2540	CTGTGGTTTCAGGTGCTGTGTGATGCATGCATGCACGACTGTGAAGACTCGGGA---TGCTCA	29					
QY	2956	GlyThrGlnProSerCysIleAspProThrLeuThrThrCysAlaAspProGlyValPro	29					
Db	2597	AGCTTCAGGCCCAACCTGTGTGCC-----ATGAGCTGCGGTCTCCCT	26					
QY	2976	-----GlnPheGly 2978						
Db	2639	CTTCACATGACTTGGT 2656						
RESULT 9	BX671124	673 bp	RNA	linear	EST 07-MAY-2004			
LOCUS	BX671124	Sus scrofa library (scac)	Sus scrofa	cdna clone				
DEFINITION	scac00361.d.22	Sprim, mRNA sequence.						
ACCESSION	BX671124							
VERSION	BX671124.1	GI:37982337						
KEYWORDS	EST.							
SOURCE	Sus scrofa	(pig)						
ORGANISM	Sus scrofa							
REFERENCE	Bukacinska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.							
TITLE	Bonner, A., Tosser-Klopp, G., Benne, F., Cabu, C., Villegier, S.,							
JOURNAL	Soares, M., Bonaldi, F. and Harey, P.							
COMMENT	A Pig Normalised Multi-Tissue cDNA Library							
	Unpublished (2003)							
	Contact: Tosser-Klopp G							
	Genetique Animale							
	Institut National de la Recherche Agronomique							
	Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan							
	Cedex, FRANCE							
	Tel: 33 (0) 5.61.28.51.14							
	Fax: 33 (0) 5.61.28.53.08							
	Email: tosser@toulouse.inra.fr							
	Sequence cleaned of vector, adaptor and repetitions. Contact us							
	at signasupport@jouy.inra.fr to obtain the chromatogram of this							
	sequence.							
	Plate: 0036 row: d column: 22.							
FEATURES	Location/Qualifiers							
SOURCE	1..673							
	/organism="Sus scrofa"							
	/mol_type="mRNA"							
	/db_xref="taxon:9623"							

US-10-016-248-2 (1-3104) x BB309178 (1-659)

QY 876 AsnTyrAsnAsnAsnHisGluCysAlaTyrSerIleGlnThrGlnProGlyIleVal 895
 DB 60 AACTACATTAACACACAGAGTGTACTTCCATCCAGACCCGCGGAGAGGAGATT 119
 QY 896 GlnLeuValAaArgAlaPheGluLeuSerGluIleAspValLeuValTyrAspGly 915
 DB 120 CAGGTTAAGCCAGGACATTGAGCTGCTTAAGAGACCTCCCAAGGCTTAAGAGCC 179
 QY 916 AsnAsnAsnSerAlaArgLeuLeuGlyValPheSerHisSerGluMetGlyValThr 935
 DB 180 ACTAACAACTCTGCCCTTGTGCTGGGGCTCTTCAGCCGCTCTGAGATGTGGGGTACT 239
 QY 936 LeuAsnSerThrSerSerSerSerLeuTriLeuAspPheIleThrAspAlaGluAsnThrSer 955
 DB 240 TTGAACAGACAGCTTAGAGCTTGTGGCTTACTTCAACAGATGCAAGAAACACACAGC 299
 QY 956 LysGlyPheGluLeuHisPheSerSerPheGluLeuIleLysCysGluAspProGlyThr 975
 DB 300 AAAGGCTTTGAACCTTCAAGTTCTAGTTTTCAGCTCATCCAGATGAGAGACCCAGGAGC 359
 QY 976 ProLysPheGlyTyrLysValHisAspGluGlyHisPheAlaGlySerSerValSerPhe 995
 DB 360 CCACAGTTCGGCTCAAGAGTTCACGATGAGAGTCAATTTTCAGAGGAGACTCGGTGACCTTC 419
 QY 996 SerCysAspProGlyTyrTyrSerLeuArgGlySerGluLeuLeuCysLeuSerGlyGlu 1015
 DB 420 AGCTGTACCTGCTGCTCAGTCTTCTGGGGCAGTAGAAGTGTCTGAGTGTGAA 479
 QY 1016 ArgArgThrTyrAspAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValArg 1035
 DB 480 CGCAGGACGTGGAGCCAGCCGCTCCCAACCTGTGTGCTGAGTGTGAGAGGAGCAGTAAGA 539
 QY 1036 GlyGluValSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeu 1055
 DB 540 GGAAGAGTGTGAGGAGGAGGTGCTGTGCGCCGCGGTACCCGCTCCCTATGAACACAACTCTC 599
 QY 1056 AsnCysAlaIleTyrThrIleGluAlaGluAlaGlyCysThrIleGly 1070
 DB 600 AACTGCACTGTGACCTATCGAGGACAGACCGCGGAGTGCACCATTTGGG 644
 RESULT 12
 CKJ12578 707 bp mRNA linear EST 01-MAR-2004
 LOCUS SB02035A2B11.f1 normalized Keck-Tagu Library SB02 Taeniopygia
 DEFINITION guttata cDNA clone SB02035A2B11.f1 5, mRNA sequence.
 CKJ12578
 ACCESSION CKJ12578.1 GI:44822152
 VERSION EST.
 KEYWORDS
 SOURCE Taeniopygia guttata
 ORGANISM Taeniopygia guttata
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Passeriformes; Estrildidae;
 Serilidae; Taeniopygia.
 REFERENCE 1 (bases 1 to 707)
 AUTHORS Clayton, D.F., Arnold, A.P., Ball, G.F., Brenowitz, E., George, J.M.,
 Mello, C.V., Wade, J., Replogle, K., Lewin, H., Band, M., Hernandez, A.,
 and Liu, L.
 TITLE The Songbird Neurogenomics Initiative: An Evolving Public Resource
 for Study of Genes, Brain, and Behavior
 JOURNAL Unpublished (2004)
 COMMENT Contact: David F. Clayton
 University of Illinois
 B107 CLSL, 601 S. Goodwin, Urbana, IL 61801, USA
 Tel: 217 244 3668
 Fax: 217 244 1648
 Email: dclayton@uiuc.edu
 Base Calling/Quality Scores: PHRED from Washington University
 Genome Center.
 Vector Trimming: Cross_match from Washington University Genome
 Center PHAP suite. Low quality bases (Phred score < 20) were
 trimmed from both ends of the sequence by an in-house script.
 This sequence is vector free and at least 200 bp in length. Funded

by PHS grant # R01 NS045264, 'Songbird Neurogenomics Initiative.'
 PCR Primers
 FORWARD: TAATACGACTCACTATAGG (T)
 BACKWARD: ATTAACTCTACTTAAG (T)
 Insert Length: 707 Std Error: 0.00
 Plate: SB02035A2 row: B column: 11
 Seq primer: TAATACGACTCACTATAGG (T)
 High quality sequence stop: 707.
 Location/Qualifiers
 1..707
 /organism="Taeniopygia guttata"
 /mol_type="mRNA"
 /db_xref="taxon:59729"
 /clone="SB02035A2B11.f1"
 /tissue_type="brain"
 /dev_stage="late embryo, post-hatch days 1, 10, 20, 45,
 and adult (pooled)"
 /lab_host="DH10B"
 /clone_lib="normalized Keck-Tagu Library SB02"
 /notes="Organ: brain; Vector: PBS II SK(+); Site: 1;
 EcorI(5' side of insert); Site 2: NotI (3' side of
 insert); The library was constructed and normalized as
 described by Bonaldo, M.F., Lennon, G. and Soares, M.B.
 (1996), Genome Research 6(9): 791-806. An identifying tag
 was added at the 3' during cDNA synthesis:
 insertCAAAAAAAAAAAAAAAAAATGCA."
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,01e-76 Length: 707
 Score: 1053.00 Matches: 190
 Percent Similarity: 93.43% Conservative: 9
 Best Local Similarity: 89.20% Mismatches: 14
 Query Match: 6.20% Indels: 0
 DB: 7 Gaps: 0
 US-10-016-248-2 (1-3104) x CKJ12578 (1-707)
 QY 1514 AAlaIleValArgPheGluCysAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGlu 1533
 DB 3 GCNNNNNGCGCTTCAGTGCAGCCCGGCTTACCGCTGGAGGCTCCCTCCATCCAG 62
 QY 1534 CysLeuProValProGlyAlaLeuAlaGluTyrAsnValSerAlaProThrCysValVal 1553
 DB 63 TGCTTGCCACATGCCCGGGGCTTGGCAGTGGAACTCTCCGCCACCTGGTGTG 122
 QY 1554 ProCysGlyValLeuLeuThrGluArgArgGlyThrIleLeuSerProGlyPheProGlu 1573
 DB 123 CCTGCGGGGGGAACCTGACGAGAGAGAAAGAACATCTGCTCCAGGCTTCTGAG 182
 QY 1574 ProTyrLeuAsnSerLeuAsnCysValITrPylsIleValValProGluGlyAlaGlyIle 1593
 DB 183 CCTTACCTGGAACAGCCCTCAACTGGTGTGAGAGTCAAGTGCCTCCGAGGAGCAGGATC 242
 QY 1594 GlnIleGlnValValSerPheValThrGluGlnAsnTyrAspSerLeuGluValPheAsp 1613
 DB 243 CAGATCCAGGTGATCAAGTTTGTGACGAGAGACAGACTGGGATTCCTCGGAAGTGTGAT 302
 QY 1614 GlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaLeu 1633
 DB 303 GGAGAGAACAAACACAGCACCATGCTGGGGAGCTTCTGGAACACAGTGCCTGCTG 362
 QY 1634 LeuAsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSerValSerAla 1653
 DB 363 CTCAACAGCACTTCAACACAGCTCTATCTCAATTCGACATCAGCGTCTCGCT 422
 QY 1654 AlaGlyPheHisLeuGluTyrLysThrValGlyLeuSerSerCysProGluProAlaVal 1673
 DB 423 GCGGCTTTTCACTGGAAATACAAACCGTGGGCTCTCCAGCTCCCGAGACCGGCGTT 482
 QY 1674 ProSerAsnGlyValLysThrGlyGluArgTyrLeuValAsnAspValValSerPheGln 1693
 DB 483 CCTGGAAATGGGCTGAAAGATCGGGAGCGCTACTGCTGTAACGACGTGTCTTCCAG 542

QY 1694 CYSGLUProGlyTyrAlaLeuGlnGlyVHISAlaHisIleSerCysMetProGlyThrVal 1713
 DB 543 TGGAGCCCGGCTACCCCTGACGGGCACTCCACATCTCTGTGATGCCGGGCAAGGTG 602
 QY 1714 ATGATGTTPAantTyProProProLeuGlyCysIleAlaGln 1726
 DB 603 CGCCGCTGGAAATACCCCTCCACCCCTCTCATCGGTAG 641
 RESULT 13
 AL638364 661 bp mRNA linear EST 07-NOV-2003
 LOCUS AL638364 XGC-neurula Xenopus tropicalis cDNA clone TNeu018n21 5',
 DEFINITION mRNA sequence.
 ACCESSION AL638364
 VERSION AL638364.2 GI:38217218
 KEYWORDS EST.
 SOURCE Xenopus tropicalis (western clawed frog)
 ORGANISM Xenopus tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 Xenopodinae; Xenopus; Silurana.
 1 (bases 1 to 661)
 Sanger Xenopus tropicalis EST project 2001 (11_2003)
 JOURNAL Unpublished (2003)
 COMMENT On Nov 7, 2001 this sequence version replaced gi:16790343.
 CONTACT: Huckle B
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 cDNA was oligo dt primed from 5ug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
 5' end and NotI at the 3' end.
 Vector: pCS107; Site 1: EcoRI; Site 2: NotI
 Host: Escherichia coli DH10B
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE ID: TNeu018n21.plkSP6
 Sequencing primer: SP6.
 Location/Qualifiers
 1..661
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TNeu018n21"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end."
 ORIGIN
 Alignment Scores:
 Pred. No.: 8 03e-73 Length: 661
 Score: 1011.00 Matches: 174
 Percent Similarity: 88.58% Conservative: 20
 Best Local Similarity: 79.45% Mismatches: 25
 Query Match: 5.95% Indels: 0
 DB: 1 Gaps: 0
 US-10-016-248-2 (1-3104) x AL638364 (1-661)
 QY 2800 LeuAlaGluGlyAsnAspPheArgTyrAsnLysThrValThrTyrGlnCysValProGly 2819
 DB 2 CTACGCTGACACCATGATTTCCAGTACACACACAGTGTCTCTCAAGTGTGCTTGA 61
 QY 2820 TyrMetMetGluSerHisArgValSerValLeuSerCysThrLysAspArgThrTyrAsn 2839
 DB 62 TATATGATGAATTCAGACAAAGCACCCCATGTTACTGTGACCAAGACCGTACCTGAAT 121

QY 2840 GlyThrLysProValCysLysAlaLeuMetCysLysProProProLeuIleProAsnGly 2859
 DB 122 GGCACAAAACCTCTCTGCAAGAAGATCAGCTGTTCATCCACCTGTGATCCCCAATGG 181
 QY 2860 LysValValGlySerAspPheMetTyrGlySerSerValThrTyrLysCysLeuGlnGly 2879
 DB 182 AAAGCAGTGGGGTCTACTTCCCTGTGGGGTCCAGCTCAATACATCACTGTGTGGAGGT 241
 QY 2880 TyrGlnLeuSerLeuProAlaValPheThrCysGlnGlyAsnGlySerThrPheGlyGlu 2899
 DB 242 TACCAACTTCTCTCTTCCAGCTGTCTTACATGTGAAGTAAACAGGAGTGGGCGG 301
 QY 2900 LeuProGlnCysPheProValPheCysGlyAspProGlyValProSerArgLysArg 2919
 DB 302 CTTCACAAATGCTCCCTGTCTGTGTGGCGATCCCGGATCCAGCCCATCTGTAG 361
 QY 2920 GluAspArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuVal 2939
 DB 362 GTGACAGAGATTCAGTACCGCTTATCTATTTCTCTCCCTCCCTCCCTTGTG 421
 QY 2940 LeuValGlySerProArgArgPheCysGlnSerAspGlyThrTyrSerGlyThrGlnPro 2959
 DB 422 CTGTGGGTCTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 481
 QY 2960 SerCysLysAspProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyTyr 2979
 DB 482 AGCTGATTTGATCCCTCCACACACATGACCTGAGATCCCATCTTTTGATG 541
 QY 2980 GlnAsnAsnSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlyLysGly 2999
 DB 542 CAGATTAACCTACCAAGAGTTTCCAGTTTGAAGTACAGTCTGTGTGTGTGTGTGTGTGT 601
 QY 3000 TyrLeuLeuGlnGlySerThrThrArgThrCysLeuProAlaLeuThrTyrSerGly 3018
 DB 602 TACCTCTTGCAAGGTTCAACAAGAACTGTCACAATCTTACCTGAGTGA 658
 RESULT 14
 BY734600 656 bp mRNA linear EST 17-DEC-2002
 LOCUS BY734600
 DEFINITION BY734600 RIKEN full-length enriched, mammary gland RCB-0526
 Y99-MC(A) cDNA Mus musculus cDNA clone G830002124 5', mRNA
 sequence.
 ACCESSION BY734600
 VERSION BY734600.1 GI:27147727
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 656)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamakawa, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schondach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V.,
 Choitha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Maglott, D.R.,
 Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, M.J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, U.U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, U.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sanderlin, A., Schneider, C., Semple, C.A., Setou, M., Shinada, K.,
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verrardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, J., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,

/clone.lib="Sus Scrofa library (scac)"
/note="Vector: pT73D-pac vector; tissues: adipose tissue,
brain, kidney, liver, muscle, ovary, testis, heart,
hypothalamus, pancreas, skin, spleen, thymus, placenta,
pituitary gland, seminal vesicle, small intestine,
uterus, adrenals, bulbo urethral gland, cerebral trunk,
epididymis, female gonad, gall-bladder, hippocampus,
large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Alignment Scores:

Pred. No.:	2,39e-66	Length:	671
Score:	933.00	Matches:	165
Percent Similarity:	98.83%	Conservative:	4
Best Local Similarity:	96.49%	Mismatches:	2
Query Match:	5	Indels:	0
DB:	5	Gaps:	0

US-10-016-248-2 (1-3104) x BX676029 (1-671)

```
QY 1727 CysGlyGlyThrValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAsn 1746
      |||::|||
DB 3 TGTGGGGAGCAGTGGAAAGACATGAGGGGGTATCTTGACCCCGGCTTCCAGGCAAC 62
QY 1747 TyrProSerAsnMetAspCysSerTrpLysIleAlaLeuProValGlyPheGlyAlaHis 1766
      63 TACCCACCAACATGAGACTGTTCCTGGAAATAGCCCTGCCGTGGGCTTTGGAGCTCAC 122
QY 1767 IleGlnPheLeuAsnPheSerThrGluProAsnHisAspTyrIleGluIleArgAsnGly 1786
      |||::|||
DB 123 ATCCAGTTCCTGAACCTTCTCCACGAGCCCAACATGATCATCGAAATCCGGAAATGGC 182
QY 1787 ProTyrGluThrSerArgMetMetGlyArgPheSerGlySerGluLeuProSerSerLeu 1806
      183 CCTACGAGACCAAGCCGATGATGAGGAGATTACGGGAGCGAGCTGCCAGGCCCTG 242
QY 1807 LeuSerThrSerHisGluThrThrValTyrPheHisSerAspHisSerGlnAsnArgPro 1826
      |||::|||
DB 243 CTCTCCAGTCTCCACGAGACCAAGCTGTACTTCCACAGCACCATCTCCAGAACCGGCGC 302
QY 1827 GlyPheLysLeuGluTyrGlnAlaTyrGluLeuGlnGluCysProAspProGluProPhe 1846
      303 GGATTCAAGCTGGAGTACAGGCGCTATGAATTCAGAGTGCCAGACCAAGACCTTT 362
QY 1847 AlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGlnSerValThrPheGluCys 1866
      363 GCCAATGACATCGTAAAGGGGTGCTGCTACACGTTGGCAGTCTGTAACTTCGAGTGC 422
DB 1867 LeuProGlyTyrGlnLeuThrGlnLysProValLeuThrCysGlnHisGlyThrAsnArg 1886
      423 CTCCGGGGGTATCAACTGATGAGGCGACCTGTCTCATCATGTACAGCAGCAACCAACCG 482
QY 1887 AsnTrpAspHisProLeuProLysCysGluVal 1897
      |||::|||
DB 483 AACTGGAGCACCCCTGCGCCAGGTGTGAAGTG 515
```

Search completed: October 18, 2004, 15:00:36
Job time : 18322 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: October 18, 2004, 09:55:06 ; Search time 2743 Seconds

(without alignments)
5940.285 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPPPALLPCLISDCC.....RSGPVGDPSTLPSSHRSPK 3104

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 4134886 segs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO/US10016248/runat_18102004_105458_11519/app_query.fasta_1.3271
-DB=Geneseg_23Sep04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10016248 @CGN_1_1_1736 @runat_18102004_105458_11519 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseg_23Sep04: *
1: genesegq1980s: *
2: genesegq1990s: *
3: genesegq2000s: *
4: genesegq2001as: *
5: genesegq2001bs: *
6: genesegq2002as: *
7: genesegq2002bs: *
8: genesegq2003as: *
9: genesegq2003bs: *
10: genesegq2003cs: *
11: genesegq2003ds: *
12: genesegq2004s: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	16985	100.0	10136	6 ABS64375
2	16985	100.0	10136	12 ADH71165
3	16985	100.0	10136	12 ADH71137
4	16970	99.9	10136	12 ADH71167
5	16512	97.2	9951	12 ADH71141
6	16319	96.1	10655	12 ADH71135

7	15917.5	93.7	10466	12 ADH71143	Adh71143 Human gen
8	15519	91.4	12900	12 ADH71145	Adh71145 Human gen
9	14142.5	83.3	8010	6 ABS64375	ABS64375 Human cub
10	14142.5	83.3	8010	12 ADH71139	Adh71139 Human gen
11	11324	10989	12	ADH72215	Adh72215 Human gen
12	10780	63.5	10433	6 AAD33320	Aad33320 Human C3b
13	10737	63.2	10673	6 AAD33318	Aad33318 Human C3b
14	10720	63.1	12525	6 AAD33319	Aad33319 Rat C3b/C
15	9077.5	53.4	6004	6 AAL49944	Aal49944 Human mol
16	7643.5	45.0	7323	6 AAS18805	Aas18805 DNA encod
17	7643.5	45.0	8034	6 AAS18806	Aas18806 DNA encod
18	7610.5	44.8	6409	6 AAS18803	Aas18803 DNA encod
19	7159	42.1	5598	6 AAS18802	Aas18802 DNA encod
20	6991	41.2	6145	6 AAS18802	Aas18802 DNA encod
21	6978	41.1	5667	6 AAS18804	Aas18804 DNA encod
22	5172.5	30.5	3896	6 AAF87127	Aaf87127 NOV16 cod
23	5172.5	30.5	3896	12 ADH71163	Adh71163 Human gen
24	5129	30.2	3904	12 ADH71161	Adh71161 Human gen
25	5129	30.2	3905	4 AAF87126	Aaf87126 NOV15 cod
26	4864	28.6	2867	10 ADC30517	Adc30517 Human nov
27	4692	27.6	2607	12 ADH71153	Adh71153 Human gen
28	4268	25.1	2328	12 ADH71149	Adh71149 Human gen
29	4257	25.1	2329	12 ADH71147	Adh71147 Human gen
30	4148.5	24.4	4506	6 ABS58378	ABS58378 Protein m
31	3942	23.2	2487	4 AAS00146	Aas00146 Human CDN
32	3856.5	22.7	2139	12 ADH71151	Adh71151 Human gen
33	3239	19.1	1792	12 ADH71157	Adh71157 Human gen
34	2894.5	16.8	3019	12 ADH72219	Adh72219 Human gen
35	2694	15.9	2609	6 AAD37600	Aad37600 Human int
36	2685	15.8	2026	10 ADC30413	Adc30413 Human nov
37	2639	15.5	2387	6 ABN99361	Abn99361 Human sec
38	2025.5	11.9	1749	10 ADC30885	Adc30885 Human nov
39	1940.5	11.4	1615	6 ABA00063	Abao0063 CAHRP-10
40	1852	10.9	11230	6 AAD32026	Aad32026 Mouse C3b
41	1797.5	10.6	10878	6 AAD32025	Aad32025 Human C3b
42	1796.5	10.6	11158	6 ABT08490	Abt08490 Human nov
43	1796.5	10.6	11158	12 ADH71299	Adh71299 Human gen
44	1796.5	10.6	11158	12 ADH71301	Adh71301 Human gen
45	1796.5	10.6	11158	12 ADH71303	Adh71303 Human gen

ALIGNMENTS

RESULT 1

ABS64375 standard; DNA, 10136 BP.

ABS64375;

15-NOV-2002 (first entry)

Human cub and sushi domain containing gene #1.

KW Human; NOVX; neurodegenerative disease; Alzheimer's disease; anxiety;
KW Parkinson's disease; Huntington's disease; neurological disorder;
KW echizophrenia; manic depression; mental retardation; angina pectoris;
KW cardiovascular disease; acute heart failure; myocardial infarction;
KW muscular disease; muscular disorder; retinal disease; photoreception;
KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;
KW immunological disorder; inflammatory disorder; immune disease; diabetes;
KW bacterial infection; fungal infection; protozoal infection; obesity;
KW viral infection; reproductive system disorder; metabolic disturbance;
KW anorexia; wasting disorder; chronic disease; infectious disease;
KW dyslipidaemia; cub; sushi; myelin; von willebrand factor; kiellin;
KW semaphorin; serine/threonine protein kinase; TGF-beta binding;
KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;
KW tolloid-like 2; cysteine sulfinic acid decarboxylase; gene; ds;
KW single nucleotide polymorphism; SNP.

XX Homo sapiens.
XX Key Location/Qualifiers
XX Variation replace(3082,A)

```

FT FT /tag= a
FT FT /note= "Single nucleotide polymorphism (SNP)"
FT FT replace(3120,A)
FT FT /tag= b
FT FT /note= "Single nucleotide polymorphism (SNP)"
FT FT replace(3251,C)
FT FT /tag= c
FT FT /note= "Single nucleotide polymorphism (SNP)"
FT FT replace(4085,G)
FT FT /tag= d
FT FT /note= "Single nucleotide polymorphism (SNP)"
FT FT /note= "Single nucleotide polymorphism (SNP)"
XX PN WO200264791-A2.
XX XX
XX PD 22-AUG-2002.
XX PF 10-DEC-2001; 2001WO-US048369.
XX PR 08-DEC-2000; 2000US-0254329P.
XX PR 14-DEC-2000; 2000US-0255648P.
XX PR 15-MAY-2001; 2001US-0291037P.
XX PR 08-JUN-2001; 2001US-02971173P.
XX PR 08-JUN-2001; 2001US-0309258P.
XX PR 29-AUG-2001; 2001US-0315639P.
XX PR 01-OCT-2001; 2001US-0326393P.
XX PA (CURA-) CURAGEN CORP.
XX PI Alebrook JP, Anderson DW, Burgess CE, Boldog FL, Casman SJ,
XX PI Colman SD, Edinger SR, Eilerman K, Gerlach V, Gorman L, Grosse WM,
XX PI Guo X, Herrmann JL, Kekuda R, Lepley DW, Li L, MacDougall JR,
XX PI Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkens RA,
XX PI Smithson G, Spytek KA, Stone DJ, Tchervet VT, Vernet CM, Voss EZ,
XX PI Zehusen BD, Zhong H, Zhong M;
XX WPI: 2002-643486/69.
XX DR P-RSDB: ABG79168.
XX DR
XX PT New NOVX polypeptides and polynucleotides useful for treating or
XX PT preventing e.g. neurodegenerative diseases, neurological disorders,
XX PT cardiovascular diseases, muscular diseases and disorders, or
XX PT immunological diseases.
XX PS Claim 9; Page 10-12; 299pp; English.
XX XX
XX CC The present invention relates to new NOVX polypeptides. The polypeptides,
XX CC polynucleotides and antibodies are useful in the manufacture of a
XX CC medicament for treating or preventing neurodegenerative diseases (e.g.
XX CC Alzheimer's disease, Parkinson's disease, or Huntington's disease),
XX CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or
XX CC mental retardation), cardiovascular disease (e.g. acute heart failure,
XX CC angina pectoris or myocardial infarction), muscular diseases and
XX CC disorders, retinal diseases (including those involving photoreception,
XX CC deafness and keratinisation disorders), cancer (e.g. ovarian cancer or
XX CC melanoma), immunological disorders, inflammatory and immune disease,
XX CC bacterial, fungal, protozoal and viral infections, and reproductive
XX CC system disorders. The proteins of the invention may be used to screen
XX CC drugs or compounds that modulate the NOVX protein activity or expression,
XX CC as well as to treat disorders characterised by insufficient or excessive
XX CC production of NOVX protein or protein forms that have decreased or
XX CC aberrant activity compared to NOVX wild type protein, such as diabetes,
XX CC obesity, metabolic disturbances associated with obesity, anorexia and
XX CC infectious diseases and various dyslipidaemias. The nucleic acid
XX CC sequences of the invention may be used in chromosome mapping, identifying
XX CC an individual from minute biological samples (tissue typing), and in
XX CC forensic identification of a biological sample. The present nucleic acid
XX CC sequence encodes a NOVX protein of the invention
XX SO Sequence 10136 BP; 2147 A; 3060 C; 2744 G; 2185 T; 0 U; 0 Other;
Alignment Scores: 0 Length: 10136
Pred. No.:

```

```

Score: 16985.00 Matches: 3104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-016-248-2 (1-1104) x ABS64375 (1-10136)
QY 1 MetAlaGlyAlaProProAlaLeuLeuLeuProCysSerLeuIleSerAspCysCys 20
Db 1 ATGGGGGGGGCCCTCCCGCCCTTGGCTGCTGCTTGAATTCACAGCTGCTGT 60
QY 21 AlaSerArgGlnArgHisSerValGlyValGlyProSerGluLeuValLysGlnIle 40
Db 61 GCTAGCAATCAGCAGACACTCCGTCGGCGTAGAGACCTCGAGCTAGTCAAGAACAAATT 120
QY 41 GluLeuLysSerArgGlyValLysLeuMetProSerLysAspAsnSerGlnLysThrSer 60
Db 121 GAGCTGAAGTCTCGAGGTGTGAAGCTGATGCCCAAGAACAGACAGCCAGAACGCTT 180
QY 61 ValLeuThrGlnValGlyValSerGlnGlyHisAsnMetCysProAspProGlyLysPro 80
Db 181 GTGTTAACTCAGGTTGGTGTGTCACAGAGACATATATGTGTCCAGACCTTGCCATACC 240
QY 81 GluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
Db 241 GAAAGGGGCAAAAGACTAGGCTCGATTTCAGGTTAGATCCAGGCTCAGCTTCACTTGC 300
QY 101 AsnGlnGlyTyrAspLeuGlnGlySerLysArgLysIleThrCysMetLysValSerAspMet 120
Db 301 AACGAGGGCTATGACTCCAGAGGGGTCCAGAGGATCACTTATGAAGTACAGCAATG 360
QY 121 PheAlaIleTyrSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisLeu 140
Db 361 TTTGGGGCTCGAGAGCACAGGCTCAGTCCGAGCCGATGTGTGATGCCACCTT 420
QY 141 ArgGlyProSerGlyLysIleThrSerProAsnMetPhePheGlnTyrAspAsnAsnAla 160
Db 421 CGAGGCCCCCTGGGCATATATACCTCCCAATTTCCCATTCAGTATGACAAACAATGCA 480
QY 161 HisCysValTyrIleIleThrAlaLeuAsnProSerLysValLysLysLeuAlaPheGlu 180
Db 481 CACTGTGTGATCATCATCAGACACTCAACCTCCCAAGGTGATCAAGCTGCTTGGAG 540
QY 181 GluPheAspLeuGluArgGlyTyrAspThrLeuThrValGlyAspGlyGlyGlnAspGly 200
Db 541 GAGTTGATTTGGAGAGGGGCTATGACACCTGACGGTGTGTGTGTGTCAGAGATGG 600
QY 201 AspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSerProHisThrPro 220
Db 601 GACCAAGAACAGTCTCTCATATCTCCAAATGCTTCAGTACAGCCCTCACACCCCA 660
QY 221 GlySerArgLysProGluSerMetSerGlyAspIleTyrArgGlnLysTyrThrValLeu 240
Db 661 GGCTCTGCATCCAGAGACATGCTCGGGGACATCTGAGGACGAGAATGAGACTGACTT 720
QY 241 GluIleCysArgAspIleSerSerSerAspAlaArgSerLysSerValArgLysSerPro 260
Db 721 GAGTCTGTGTGATCATTTAGAGTTCAGATGACAGGTCAGTTCAGAGAGAGTCTCCA 780
QY 261 LysThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlnGlnLysSerCys 280
Db 781 AAGACTTCTAATGCTGTGGAACTTTGTTGCTCGGGAGACAGATTCAGAGGCAAGTTC 840
QY 281 GlyAspProGlyLysProAlaTyrGlyArgArgGlnGlySerArgPheHisHisGlyAsp 300
Db 841 GGTGACCTTGACATCTGATGATGCGCGGAGGAAAGGCTCCCGGTTTCCACCGGTGAC 900
QY 301 ThrLeuLysPheGlnCysGlnProAlaPheGlnLeuValGlyGlnLysValIleThrCys 320
Db 901 ACACTCAAGTTTAAAGTCCAGCCCGCTTTGAGCTGTGAGACAAAGGCAATCAATG 960
QY 321 GlnLysAsnAsnGlnTyrSerAlaLysLysProGlyCysValPheSerCysPhePheAsn 340

```

```
Db 961 CAAAAGATACCAATGGTGGCTTAAGAGCCAGGCTGGCTTCTCTGCTTCTTCAAC 1020
Qy 341 PheTherProserGlyValValLeuSerProAenTyProGluAerTyGlyAAsnHis 360
Db 1021 TTCCACGAGCCGCTGGGGTGTCTCTCTCCCAATCAACGAGGACTTAAGGCAACAC 1080
Qy 361 LeuHisCyValTTrPleuIleLeuAlaArpProGluSerArgIleHisLeuAlaPheAsn 380
Db 1081 CTCACAGTGTCTGTGGCTCATCTTGGCCAGGCTTGAAGCCGCAATCCATCGGCTTCAAC 1140
Qy 381 AspIleAspValGluProGluInPheAspPheLeuValIleuAspGlyAlaThrAlaGlu 400
Db 1141 GACATTAAGCGTGAAGCCCTCAATTGATTCTGTGCATCAAGATGGGGCCACCGCGAG 1200
Qy 401 AlaProValLeuGlyThrPheSerGlyAAsnGluLeuProSerSerIleThrSerSerGly 420
Db 1201 GCGCCCTCTCCGGGACCTTCTCAAGAAACAGCTTCTCTCCATCACAAGAGATGGC 1260
Qy 421 HisValAlaArgLeuGluPheGluThrPheHisSerThrglyAAsnGlyPheAsnIle 440
Db 1261 CAGCTGGCCCTCTCAAGTTCAGACTGACCACTCCACAGGAAAGAGGGCTTCAACATC 1320
Qy 441 ThrPheThrThrPheArgHisAsnGlyCyAspProAspProGlyValProValAsnGlyAs 460
Db 1321 ACTTTTACCACTTCCGACACAGAGTGGCCGATCCTGGCGTTCCAGTAAATGGGAAA 1380
Qy 461 ArgPheGlyAspSerLeuGluInLeuGlySerSerIleSerPheLeuCyAspGluGlyPhe 480
Db 1381 CGGTTTGGGGACACCTTCCAGCTGGGCGCTCCATCTCTCTCTGTGAAGAGGGCTTC 1440
Qy 481 LeuGlyThrGlnGlySerGluThrIleThrCyValLeuGlyGluGlySerValValTTrp 500
Db 1441 CTGGGACTCAGGGCTCAGAGACCAATCACTTGGCTCTGAAGAGAGGGAGGTGTGG 1500
Qy 501 AsnSerAlaValLeuArgCyGluValAProCyGlyGlyHisLeuThrSerProSerGly 520
Db 1501 AACAGCCGTGTGCGGGCTGAGAGCTCCCTGTGGTGTCACTGAGCTGCCACAGGGGC 1560
Qy 521 ThrIleLeuSerProGlyTTrpProGlyPheThrglyAAsnAlaLeuSerCyValTTrpVal 540
Db 1561 ACCATCTCTCTCCGGGCTGGCTGTCTTCAAGAGATGCTTGAAGTGTGGCTGGGTG 1620
Qy 541 IleGluAlaGlnProGlyTTrpProIleuSerIleThrPheAspArgPheIleThrGluVal 560
Db 1621 ATTAAAGCCAGCCAGGCTACCCCAATCAAAATCACTTGCAGCAATTCAAAACGAGGTC 1680
Qy 561 AsnTyAspThrLeuGluValArgAspGlyArgThrglySerAlaProLeuIleGlyVal 580
Db 1681 AACTATGACACCCGGAAGTACGGGATGGGGGAGCTTACAGGGCCCTTGAATGGGGTT 1740
Qy 581 TyHisGlyThrGlnValProGlnPheLeuIleSerThrSerAenTyLeuTyLeuLeu 600
Db 1741 TACACCGGAGCCAGGTTCCCACTTCTCATCAGACACAGCACTACTTCACTCTC 1800
Qy 601 PheSerThrAspLeuSerHisSerAspIleGlyPheGluLeuArgTyGluThrIleThr 620
Db 1801 TTCTCTACCGACAAGACTCACTCGACATCGGCTTCCAGCTCCGCTTGAAGACTATACA 1860
Qy 621 LeuGlnSerAspHisCyLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn 640
Db 1861 CTGCAAGTCAAGCACTGTCTGATCCAGGAATCCCAAGTAAATGACAGCGTCAAGGAAAT 1920
Qy 641 AspPheTyValGlyAlaLeuValThrPheSerCyAspSerGlyTyThrLeuSerAsp 660
Db 1921 GACTTCTAAGTGGGCGGCTGTGATCACTTCACTGTGACTCGGGCTACACTTAAGTAC 1980
Qy 661 GlnValProLeuGluCyGluProAspPheGluInTrpSerArgAlaLeuProSerCyGlu 680
Db 1981 GGGAGGCTCTGAGAGTGAAGCCCACTTCAAGTGAAGCCGGGCGCTGCGCAAGTTGGA 2040
Qy 681 AlaLeuCyGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
Db 2041 GCTCTGTGTGGCTTCAATCAAGGCTCCAGTGGAGCAATCTTGTGCGCAGGGTCCCT 2100
Qy 701 AspPheTyProAsnLeuAsnCyValThrIleIleGluThrSerHisGlyValGly 720
Db 2101 GACTTCAACCCCAACCACTTGAAGTCACTGATTAATCAAAATCTCAATGCAAGGCT 2160
Qy 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyLeuLeuIleThr 740
Db 2161 GTGTCTTCACTTCCACACTTCCACTGGAAGAGTGGCATGACTTCTCTCATCACT 2220
Qy 741 GluAsnGlySerPheThrGlnProLeuArgGluLeuThrglySerAlaLeuProAlaPro 760
Db 2221 GAGAACGGCAGCTTCAACCCAGCCCTGAGGACAGTAACTGATCTCGGCTGCCAGCTCCC 2280
Qy 761 IleSerAlaGlyLeuTyArgIleAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
Db 2281 ATCAGCGCTGGGCTTAATGCAACTTCAAGCCAGGCTGGCTTCACTCTGATTTCTC 2340
Qy 781 MetSerTyArgGluPheAsnIleThrPheSerGlyTyAspLeuGluProCyGluGlu 800
Db 2341 ATGTCAATGAAGATTCAACATCACTTCTCAAGATCAAGACTTGGAGCCCTGAGAGAG 2400
Qy 801 ProGluValProAlaTySerIleArgGlyLeuGlnPheGlyValGlyAspThrLeu 820
Db 2401 CCGAGAGTCCAGGCTACAGATCCGGAGAGGCTTGAGTTGGCGTGGCGACACCTTG 2460
Qy 821 ThrPheSerCyAspProGlyTTrpArgLeuGluGlyThrAlaArgIleThrCyLeuGly 840
Db 2461 ACCTTCTCTGCTTCCCGGGTACCTGTGAAGGACCGCCGCAATCAGCTGTGGG 2520
Qy 841 GlyArgArgLeuThrPheSerProLeuProArgCyValAlaGluCyGlyAAsnSer 860
Db 2521 GCGAGAGGCGGCTGTGAGACTCCCTCTGCAAGTGTCTGTGAGTGGGAATTC 2580
Qy 861 ValThrglyThrGlnGlyThrLeuLeuSerProAspPheProValAenTyAsnAsn 880
Db 2581 GTCAAGAGCACTCAGGGTACTTGTGTCTGCCCACTTCTGTGAATCAATTAACAT 2640
Qy 881 HisGluCyValIleTySerIleGlnThrGlnProGlyValGlyIleGlnLeuValAlaArg 900
Db 2641 CATGAATGATCTCACTCAATCCAGACCCAGGAGGAAAGGAATCGCTGAAGAGCCAGG 2700
Qy 901 AlaPheGluLeuSerGluGlyAspValLeuValTyAspGlyAAsnAsnSerAla 920
Db 2701 GCATTCGAATCTCCAGAGAGATGTCTCAAGTTTATATATGCAACCACTCCGCC 2760
Qy 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerThrSer 940
Db 2761 CGTTGTGGAGTTTATACCATTTGAGATGAGGGGTGACTTGAACAGACATCC 2820
Qy 941 SerSerLeuThrPheLeuAspPheIleThrAspAlaGluAsnThrSerIleGlyPheGluLeu 960
Db 2821 AGCAGTGTGGCTGATTCATCACTGATCACTGAAGAAACAGCAAGGGCTTTGAACGTG 2880
Qy 961 HisPheSerSerPheGluLeuIleTyCyGlyAspProGlyThrProLysPheGlyTy 980
Db 2881 CACTTTTCCAGCTTGAACATCAATATGAGAGACCCAGAAACCCCAAGTTGGGTAC 2940
Qy 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCyAspProGly 1000
Db 2941 AAGGTCATGATGAAGTCAATTTTGAGGAGAGCTCGGTCTTCACTGACCTCGGA 3000
Qy 1001 TySerLeuArgGlySerGluGluLeuLeuCyLeuSerGlyValAlaGlyGlnThrTrpAsp 1020
Db 3001 TACAGCTGTGGGTATGAGAGCTGTGTGTGATGAGAGCCCGAGACTGGGAC 3060
Qy 1021 ArgProLeuProThrCyValAlaGluCyGlyGlyThrValArgGlyGluValSerGly 1040
Db 3061 CGGCTTGGCCCACTGTGTGCGGAGTGTGAGAGGACATGAGAGAGAGTGTGGGG 3120
Qy 1041 GlnValLeuSerProGlyTyProAlaProTyGluHisAsnLeuAsnCyValIleTTrpThr 1060
Db 3121 CAGGTGTGTACCGGGGTATCCAGCTCTTAATGAACAACTCAATGATCTGTGACC 3180
```


QY	1061	ILEGLULIAGLULIAGLVCYSITR11LEGLYLEUHIAPHELEUVALPHEASPTTHGLUGLU	1080
DB	3181	ATCGAACAAGAGCGCGGCTGCACCATTTGGGCTTACCTTCCTGGTGTGACACAGAGAG	3240
QY	1081	VALHISAPVALLEUATRG1LETTPASPGYPROVALG1USEG1YVALLEULEULYGLU	1100
DB	3241	GTTACAGACGTGCTGGCCACTTGCGATGGGCTGTGGAGAGGGGGTCTGTGAAGAG	3300
QY	1101	LEUSERG1YPROVALALEUPROLYSAPLEUHISSERTHPHEASNSERVALVALLUGLU	1120
DB	3301	CTGAGTGGCGCGCGCTGGCCCAAGACCTGCATACACTTCACTGCGTGTCTCGAC	3360
QY	1121	PHESERTHPASPHETHERSERLYSGINGLYPHEALALEGINPHESERVALSERTHR	1140
DB	3361	TTTCAGCACTGACTTCTTCCACAGCAAGAGGGCTTGGCATTCATTAATTTACAGTCTCAC	3420
QY	1141	ALATHSERCYASNAAPPROG1Y1LEPROGINAENG1YSERATGSRERG1YASPERTP	1160
DB	3421	GCAACGTCTCGAATGACCCCTGGATCCCGAGAAITGGAGTCCGAGTGGTACAGTTGG	3480
QY	1161	GLUALG1YASPSERTHRTVALPHEGLNCYASAPPROG1YTRALALEUNG1YSERALA	1180
DB	3481	GAAAGCCGGAGCTCCACAGATGTTCCAGTGTGACCTGTGCTACGGCGCAGAGGAATGCA	3540
QY	1181	GLUL1SESCYAVALLYE1LEGLULANARGPHEPETHRGINPROSERPROPTHYRS	1200
DB	3541	GAGATCAGCTGTGTGAAGATCGAACAAGTTCCTTCTGGACCCCAACCCCGCAACATCC	3600
QY	1201	1LEAL1PROCYG1YGLYASPLEUTHRG1YPROSERG1YVAL1LEUSERPROANTYR	1220
DB	3601	ATGCGTCCCTGGGGGAGACCTGCACAGAACCATCTGAGTCATCTCTCAACCAATTAC	3660
QY	1221	PROGLUPROTYRPROPROG1YLEGLUCYASAPTRHLYSERVALPHEVALSERPROAPTYR	1240
DB	3661	CCAGAACCTTACCCCGACGCAAGAGTGTACCTGGAAGACCGGTCTCACCAAGACTAC	3720
QY	1241	VAL11LEALALEUVALPHEASNL1EPHEASNL1EUG1UPROG1YLYRASPHELEUHI1E	1260
DB	3721	GTTCATCGCCCTGGTATTTAAATCATCTTAACTGGAGCTGGCATGACTTCTCCATATC	3780
QY	1261	TYRASPGLYRARGASPSERLEUSERPROLEU11LEGLYSEPTERYGLYSEGLINLEUPRO	1280
DB	3781	TACGACGAGCGGAGACTCTCTACGCCCTCTCATAGAAAGCTTCATAGGCTCCACGCTCCA	3840
QY	1281	GLYARGL1EGLUSERSERASNSERLEUPHELEU1APHEATGSRASPA1ASERVAL	1300
DB	3841	GGCCGCATTTGAAGAGCAGCAACAGCTCTCTCCGCTCCGACGACATTCATCTGG	3900
QY	1301	SERASNA1AGLYPHEVAL11LEASPTYRTHGLUASNPROARG1USERCYASPHASPRO	1320
DB	3901	AGCAATGCTGGCTTGTGATTTGACTATACAGAAAACCCCGGGAGTCATGTTTATCTCT	3960
QY	1321	GLYSERT1LEYASANG1YTHRARGVALG1YSEASPLEU1LYLEUG1YSESERVALTHR	1340
DB	3961	GGTTCATACAAAGAACGCAACCGGGTGGGCTCCGACCTGAAGTGGGCTCTCCCTCACCC	4020
QY	1341	TYRTRYCYSH1EG1YGLYTRYGLUVALG1UG1YTHRSERTHRLEUSERCY11LEUG1Y	1360
DB	4021	TACTACTGCCACGGGGGCTACGAAGTTGAGGCACTGCACCCCTGAGCTGATCCTGGGG	4080
QY	1361	PROASPG1YASPROVAL1TRPANAANPROARGPROVALCYETHALAPROCYEG1Y	1380
DB	4081	CCTGATGGGAAGCCCGGTGGAACAATCCCGGCCAGTCTGCACAGCCCGCTGTGGGGGA	4140
QY	1381	GLINTYVALG1YSEASPG1YVAL1VALLEUSERPROANTYRPROGINANTYRTHRSER	1400
DB	4141	CAGATATGGGTTCCGACGAGTGTCTGTCTCCCAACTACCCCAAGACTACCAAGT	4200
QY	1401	GLYGLIN1LEYCYSEUPTYRPHENVAL1THRVALPROLYASPTYRVALPHEGL1YGLN1PHE	1420
DB	4201	GGACAGATGTGCTTGATATTTTGTACTGTGCCCCAAGACATAGTGTGTGTGGCCAGTTC	4260
QY	1421	ALAPHEPHEH1STHRTALALEUANAPVALVALG1UVALH1H1EASPG1YH1SSERTGNH1S	1440
DB	4261	GCCTTCTTACACAGCGCCCTCAACAGCATGTGGAGGTTCACAGAGGCCACAGCCAGCAC	4320
QY	1441	SERARGLEUSERSERLEUSERG1YSERH1STHRTG1YGLUSERLEUPROLEU1ATHR	1460
DB	4321	TGCGGCTCTCCAGCTCCCTCTCGGGCTCCCATACAGGAGATCATCTGCCCTTGGCACCC	4380
QY	1461	SERASNL1VALLEU1LEYSPHESER1ALYVSG1YLEU1ALPROVALARG1Y1PHEH1S	1480
DB	4381	TCCATATATTTCTCATTTAGTTACAGCGCAAAAGGCTTCGACCAAGCCAGAGGCTTCAC	4440
QY	1481	PHEVALTRYGLNALVALPROARGTHRSER1ATTRG1NCYSSERSEVALPROGLUPRO	1500
DB	4441	TTTGTTACCAAGCGGTTCTCGAACACAGCGCACAGATGCTGATCTGTGCCAAACC	4500
QY	1501	ARGTRYGLYLYEARGLEUG1YSEAPPHESERVALG1YAL1ALEVALARGPHEGLUCYS	1520
DB	4501	CGCTATGGCAAGAGCTGGGCAAGTACCTTCTGGTGGGGCCATCTGCTCCGTTGAAATGC	4560
QY	1521	ASNSERG1YTRALALEUNG1YSEPROG1Y1LEGLUCY1LEUPROVALPROGLYALA	1540
DB	4561	AACCTCGGCTATGCCCTCGAGGAGTCCGCAAGATCGAGTCTCTGTGCTGGGGCC	4620
QY	1541	LEUALGLINTTPASNA1SER1ABPROTHRCYASVALVALPROCYG1YGLYASNL1EUTHR	1560
DB	4621	TTGGCCCAATGGAATGCTCAGCGCCACAGTGTGGTGGCGGTGTGAGGCAACCTCAC	4680
QY	1561	GLUARGARG1YTHR11LEUSERPROG1Y1PHEPROG1UPROTYRLEUANSERTLEUAN	1580
DB	4681	GAGCGAGGGGACCAATCTGTCTCCCTGGTCTCCAAAGCGGTACTCAACACTTCAC	4740
QY	1581	CYVAL1TRPVAL1EVALVALPROGLUG1YALAG1Y1LEGLIN1LEGLINVAL1SERPHE	1600
DB	4741	TGTGTGTGGAAGATGTGTGTCCCGAAGCGCTGGATCCAGATCCAAATGTTCAGTTT	4800
QY	1601	VAL1THRG1UG1NASTTPASPSERLEUG1UVALPHEASPG1YAL1APASNTHRVAL1THR	1620
DB	4801	GTGACAGAGCAACTGGAGACTCCCTGGAAGTATTTGATGTGCAGATTAACATCTGTAACC	4860
QY	1621	METLEUG1YSEPHESERG1YTHRTHVALPRO1ALEUEUANSETRHRSERANG1N	1640
DB	4861	ATGCTGGGGAAGTTTCTACGAACAACCTGTCCCTCTTGAAACGACCTTCAACCG	4920
QY	1641	LEUTYRLEUHI1SPHERY1SERASPL1SERVALSER1AL1AG1Y1PHEH1SLEUG1YTR	1660
DB	4921	CTCTACCTTCAATTTCTACTCAGATATCAGGATATGCGAGCTGCTTCACCTTGAGTAC	4980
QY	1661	LYSTRVALG1YLEUSERSERCYASPROG1UPROVALVALPROSERASNL1YVAL1YSTR	1680
DB	4981	AAAAAGTGGGCTCGAGCATTTGTCCGGAACCTGTGTGCCCAAGTAAACGGGGTGAAGACT	5040
QY	1681	GLYGLUARTRYRLEUVAL1ASNAAPVAL1ASERPHEN1CYSEGL1UPROG1YTRYR1ALEU	1700
DB	5041	GGCGAGCGCTACTTGTGAATGATGTGTGTCTTCCAGTGTGAGCGGGATATGCCCTTC	5100
QY	1701		

```
Db 5341 ATGAAATCCGGAATGCCCCCTATGAGACAGCCGATGAGGAATTCAGTGAAGC 5400
Qy 1801 G1uLeuProSerSerLeuLeuSerThrsSerHsg1uThrThrValTyPheHisSerAsp 1820
Db 5401 GAGCTTCAGACTCCCTCTCTCCACGTCACGAGACACCGTGTATTTCCACAGCCAC 5460.
Qy 1821 HisSerG1naAsnArgProG1yPheLeuLeuG1uTyrg1na1aTyrg1uLeuG1uLeu 1840
Db 5461 CACTCCCAAGATCGGCCAGATTCAGCTGAGATTCAGGCTTATGAATTCAGAGATGC 5520
Qy 1841 ProAspProG1uProPheAlaAsnG1y1LeValArg1yAlaG1yTyPheAsnValG1yG1n 1860
Db 5521 CCAGACCCAGAGCCCTTTGCCAATGGCATTTGAGAGGAGCTGGCTTACAGTGGAGCA 5580
Qy 1861 SerValThrPheG1uCyLeuLeuProG1yTyrg1nLeuThrg1yHisProValLeuThrCys 1880
Db 5581 TCAGTGAACCTTCGATGCTCTCCCGGGGTATCAATTGACCTGGCCACCTGTCTCACTGT 5640
Qy 1881 G1nH1sg1yThrAsnArgAsnTrpAspHisProLeuProCysG1uValProCysG1y 1900
Db 5641 CAACATGGCACCAACCGGAACCTGGGACCAACCCCTGGCCAGATGTAAGTCCCTGTGAGC 5700
Qy 1901 G1yAsn1LeThrSerSerAsnG1yThrValTyPheProG1yPheProSerProTyPheSer 1920
Db 5701 GGGAACATCACTTCTTCCACAGGCACTGTATCTCCGGGGTTCCTAGCCCTGATCC 5760
Qy 1921 SerSerG1naAspCyVal1TrpLeu1LeThrValPro11eg1yHsg1yValArgLeuAsn 1940
Db 5761 AGCTCCCGAGACTGTGTGCTGCTGATCAACCGGCCATTTGGCCATGGCGGTCTCTCAAC 5820
Qy 1941 LeuSerLeuLeuG1nThrg1uProSerG1yAspPhe1LeThr1LeTrpAspG1yProG1n 1960
Db 5821 CTCAGCTGCTGACAGACAGACCTCTGTGAGATTTCAACCATCTGGAGATGGGCCACAG 5880
Qy 1961 G1nThrAlaProArgLeuG1yValPheThrArgSerMetAla1yValTyPheValG1nSer 1980
Db 5881 CAACAGACCAACCGGCTGGGCTTTCACCCGGACATGGCCAGAAACAGTGAAGT 5940
Qy 1981 SerSerAsnG1nValLeuLeuLeuPheHisArgAspAla1aThrG1yG1y1LePheAla 2000
Db 5941 TCATCCCAACAGGCTCTCTCAAGTTCACCGTGTATGACGACACAGGGGGATCTTCCCC 6000
Qy 2001 1LeAlaPheSerAlaTyPheLeuThrTyPheCysProProThr1LeLeuProAsnAla 2020
Db 6001 ATAGCTTCTCGCTTATCTCACTCACCAATGCCCTCTCCACCATCTCCCAAGCC 6060
Qy 2021 G1uValValThrg1uAsnG1uG1uPheAsn11eg1yAsp11eValArgTyPheArgCysLeu 2040
Db 6061 GAAGTCGTCAAGAGATGAAGAAATTCATATAGTGAATCGTACGCTTACAGATGCTC 6120
Qy 2041 ProG1yPheThrLeuValG1yAsnG1u1LeLeuThrCysTyPheLeuG1yThrTyPheLeu 2060
Db 6121 CCTGGCTTACCTTATGAGGGGAATGAATTTCTGACCTGCAAACTTGAACCTTACCTGAG 6180
Qy 2061 PheG1uG1yProProPro11eCyG1uValHisCysProThrAsnG1uLeuLeuThrAsp 2080
Db 6181 TTGAGGAGCAACCCCGATATGTGAATGACGTCTCCAAACAAGACCTTCTGACAGAC 6240
Qy 2081 SerThrg1yVal11eLeuSerG1nSerTyPheProG1ySerTyPheProG1nPhG1nThrCys 2100
Db 6241 TCCACAGGCGTGAATCTCTAGCCAGACGATACCTCTGAACTATCCCACTTCCAGACCTGC 6300
Qy 2101 SerTrpLeuValArgValG1uProAspTyPheAsn11eSerLeuThrValG1uTyPheLeu 2120
Db 6301 TCTTGCTGTGTGAGGTGAGGCCGACATATAACATCTCCCTCACAGTGAATCACTTCTC 6360
Qy 2121 SerG1yG1yG1nTyPheAspG1uPheG1u1LePheAspG1yProSerG1yG1nSerProLeu 2140
Db 6361 AGCGAAGACATATGATGATGATTTTGTATGATTCATCAAGCAAGTCTCTG 6420
Qy 2141 LeuTyPheAlaLeuSerG1yAsnTyPheSerAlaProLeu11eValThrSerSerAsnSer 2160
Db 6421 CTGAAGCCCTCAGTGGGAATTAATTAATCACTCCCTGATTTGTCACAGCTCAAGAACTCT 6480
Qy 2161 ValTyPheLeuArgTrpSerSerAspHisAlaTyPheAsnArg1yG1yPheVal1LeArgTyP 2180
Db 6481 GTGTAACCTGCGTGTATCTGATTCAGCTCAATGGAAGGGCTTCAAGATCCGCTAT 6540
Qy 2181 SerAlaProTyPheCysSerLeuProArgAlaProLeuHisG1yPhe11eLeuG1yG1nThr 2200.
Db 6541 TCAGCCCTTAATCACTGAGCTTCCAGAGGCTTCACTTCAATGGCTTCACTTACGCAAGCC 6600
Qy 2201 SerThrg1nProG1yG1ySer11eHisPheG1yCysAsnAlaG1yTyPheArgLeuValG1y 2220
Db 6601 AGCACCAACCGCGGGGCTCCATCCATTTGGTGAACGCGGCTACCGCTGTGGGA 6660
Qy 2221 HisSerMetAla11eCySerThrArgHisArgProG1nG1yTyPhe11eLeuThrPheSerG1uAla11e 2240
Db 6661 CACAGCATGGCAATCTGTATCCCGGACCCCGAGGATCAACCTGTGAGGAGCAAGCCATC 6720
Qy 2241 ProLeuCyG1nAlaLeuSerCyG1yLeuProG1uAlaProCysAsnG1yMetValPhe 2260
Db 6721 CCTCTCTGTCAGCTCTTCTCTGTGGCTTCTTGAAGCCCAAGAAATGAAATGCTGTT 6780
Qy 2261 G1yTyPheG1uTyPheThrValG1yThrTyPheAlaValTyPheCysSerG1uG1yTyPheHisLeu 2280
Db 6781 GGCAAGAGATACAGATGGGAACCAAGCCGTGTACAGCTGATGAAGGCTTACCACTC 6840
Qy 2281 G1nAlaG1yAlaG1uAlaThrAlaG1uCyLeuAspPheThrg1yLeuTrpSerAsnArgAsn 2300
Db 6841 CAGCAGAGCGGTGAGGCACTGACAGATGTGTGACACAGCCCTATGAGCAACCGCAAT 6900
Qy 2301 ValProProG1nCyValProValThrCysProAspValSerSer11eSerValG1yHis 2320
Db 6901 GTCCACCAAGTGTGCTCTGTATCTTGTATCTTGTATCTGATGATGATGATGATGAT 6960
Qy 2321 G1yArgTrpArgLeu11ePheG1uThrg1nTyG1nPhG1nAlaG1nLeuMetLeu11e 2340
Db 6961 GGCCGATGAGAGCTTATCTTTGAGACACAGATCAGTTCCAGGCCCAAGCTATCTCATC 7020
Qy 2341 CyAspAspProG1yTyPheTyPheThrg1yG1nArgVal11eArgCyG1nAlaAsnG1yCys 2360
Db 7021 TGTATCCCTGCTCTATCTATACAGGCAAGGATCTCCGTGAGGCCAAATGGCAAA 7080
Qy 2361 TrpSerLeuG1yAspSerThrProThrCysArg11e11eSerCyG1yG1uLeuPro11e 2380
Db 7081 TGAAGCTCGGGAGCTTACGCCCACTCGCAATCACTCTGTGAGAGCTCCGAT 7140
Qy 2381 ProProAsnG1yHisArg11eG1yThrLeuSerValTyrg1yAlaThrAla11ePheSer 2400
Db 7141 CCCCCAATGGCCACCGCATGGAACACTGTCTGTACGGGGCAACAGCCATCTTCTCC 7200
Qy 2401 CyAspAspG1yTyPheThrLeuValG1ySerArgVal11eG1uCyPheAlaAsnG1yLeu 2420
Db 7201 TGCATATTCGGATACACACTGTGTGGCTCCAGGGTGTGTGATGTCAGTCCAAATGGGCTC 7260
Qy 2421 TrpSerG1ySerG1uValArgCyPheLeuAlaG1yHisCysG1yThrProG1uPro11eVal 2440
Db 7261 TGAAGTGTCTGAAGTCCGTGCTGTGGAACCTGTGGAACCTGTGGAACCTGTGTC 7320
Qy 2441 AsnG1yHis11eAsnG1yG1uAsnTyPheSerTyPheArg1ySerVal1Tyrg1nCyAsn 2460
Db 7321 AAGGAGACATCAATGGGAGAACTACAGCTACGGGGAGTGTGTACCAATGCAAT 7380
Qy 2461 AlaG1yPheArgLeu11eG1yMetSerValArg11eCyG1nG1nAspHis11eTrpSer 2480
Db 7381 GCTGCTTCCGCTCGATGCGCATGTCTGTGTGCAATCTGCGACAGAGATCACTGTGTG 7440
Qy 2481 G1yTyPheThrProPheCysValPro11eThrCysG1yHisPheProG1yAsnProValAsnG1y 2500
Db 7441 GGCAGAGCCCTTCTGTGTGTCATTAATCTGTGGAACCCAGGCAACCTGTATAAGGC 7500
Qy 2501 LeuThrg1nG1yAsnG1nPhAsnLeuAsnAspVal1ValTyPheValCyAspAspProG1y 2520
Db 7501 CTCACCTAGGGTAAACAGTTTAACCTCAAGATGTGTCAAGTTGTTCAAACCTGTGG 7560
```


PR	04-JUN-2002;	2002US-0385794P.
PR	05-JUN-2002;	2002US-0386041P.
PR	05-JUN-2002;	2002US-0386047P.
PR	06-JUN-2002;	2002US-0386376P.
PR	06-JUN-2002;	2002US-0386453P.
PR	06-JUN-2002;	2002US-0386864P.
PR	07-JUN-2002;	2002US-0387016P.
PR	07-JUN-2002;	2002US-0387966P.
PR	07-JUN-2002;	2002US-0388616P.
PR	07-JUN-2002;	2002US-0389311P.
PR	07-JUN-2002;	2002US-0389642P.
PR	07-JUN-2002;	2002US-0389712P.
PR	07-JUN-2002;	2002US-0389762P.
PR	08-JUN-2002;	2002US-0389650P.
PR	10-JUN-2002;	2002US-0387400P.
PR	11-JUN-2002;	2002US-0387555P.
PR	11-JUN-2002;	2002US-0387610P.
PR	11-JUN-2002;	2002US-0387625P.
PR	11-JUN-2002;	2002US-0387634P.
PR	11-JUN-2002;	2002US-0387668P.
PR	11-JUN-2002;	2002US-0387686P.
PR	11-JUN-2002;	2002US-0387702P.
PR	11-JUN-2002;	2002US-0387836P.
PR	11-JUN-2002;	2002US-0387859P.
PR	12-JUN-2002;	2002US-0387933P.
PR	12-JUN-2002;	2002US-0387934P.
PR	12-JUN-2002;	2002US-0387960P.
PR	12-JUN-2002;	2002US-0388022P.
PR	13-JUN-2002;	2002US-0388096P.
PR	13-JUN-2002;	2002US-0389123P.
PR	14-JUN-2002;	2002US-0389118P.
PR	14-JUN-2002;	2002US-0389120P.
PR	14-JUN-2002;	2002US-0389146P.
PR	14-JUN-2002;	2002US-0389146P.
PR	17-JUN-2002;	2002US-0389729P.
PR	17-JUN-2002;	2002US-0389742P.
PR	18-JUN-2002;	2002US-0389884P.
PR	19-JUN-2002;	2002US-0390006P.
PR	19-JUN-2002;	2002US-0390209P.
PR	21-JUL-2002;	2002US-0390763P.
PR	17-JUL-2002;	2002US-0390766P.
PR	06-AUG-2002;	2002US-0401628P.
PR	09-AUG-2002;	2002US-0402156P.
PR	09-AUG-2002;	2002US-0402256P.
PR	09-AUG-2002;	2002US-0402389P.
PR	12-AUG-2002;	2002US-0402786P.
PR	12-AUG-2002;	2002US-0402816P.
PR	12-AUG-2002;	2002US-0402821P.
PR	12-AUG-2002;	2002US-0402832P.
PR	13-AUG-2002;	2002US-0403448P.
PR	13-AUG-2002;	2002US-0403459P.
PR	13-AUG-2002;	2002US-0403531P.
PR	13-AUG-2002;	2002US-0403532P.
PR	13-AUG-2002;	2002US-0403563P.
PR	13-AUG-2002;	2002US-0403617P.
PR	15-AUG-2002;	2002US-0403617P.
PR	26-AUG-2002;	2002US-0406182P.
PR	26-AUG-2002;	2002US-0406355P.
PR	27-AUG-2002;	2002US-0406240P.
PR	12-SEP-2002;	2002US-0410084P.
PR	20-SEP-2002;	2002US-0412528P.
PR	23-SEP-2002;	2002US-0412731P.
PR	30-SEP-2002;	2002US-0414801P.
PR	30-SEP-2002;	2002US-0414839P.
PR	30-SEP-2002;	2002US-0414840P.
PR	30-SEP-2002;	2002US-0414954P.
PR	09-OCT-2002;	2002US-0417406P.
PR	23-OCT-2002;	2002US-0420639P.
PR	28-OCT-2002;	2002US-0421156P.
PR	31-OCT-2002;	2002US-0422690P.
PR	01-NOV-2002;	2002US-0423130P.
PR	05-NOV-2002;	2002US-0423798P.

PR	05-NOV-2002;	2002US-0423798P.
PR	12-NOV-2002;	2002US-0425453P.
XX	(CURA-) CURAGEN CORP.	
PI	Alsebrook JP, Alvarez E, Anderson DW, Boldog FL, Caseman SJ,	
PI	Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,	
PI	Ettnerberg S, Gangoli EA, Gerlach VL, German L, Gunther E, Guo X,	
PI	Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,	
PI	MacLachlan T, Malynkar UM, Meixick AJ, Millet I, Mishra VS,	
PI	Padiagar M, Patrujan M, Pena CEA, Peyman JA, Raha D, Rastelli L,	
PI	Rieger DK, Rothenberg ME, Schore P, Shenoy SG, Shinkets RA,	
PI	Smithson G, Spytek KA, Stone DJ, Vernet CM, Voese EZ, Zhong M,	
PI	Zhong H,	
DR	WPI, 2004-081935/08.	
DR	P-PSDB; ADH71166.	
XX		
PT	New NOX polypeptides and nucleic acid molecules useful for preventing or	
PT	treating NOX-associated disorders, e.g., cancer, diabetes, infection or	
PT	obesity, and in chromosome mapping, tissue typing or pharmacogenomics.	
XX		
PS	Example 4; SEQ ID NO 61; 1880bp; English.	
CC	The invention relates to a novel isolated polypeptide (NOX). A	
CC	polypeptide of the invention has cytostatic, immunomodulator,	
CC	neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and	
CC	antihypertensive activity, and may have a use in gene therapy, and as a	
CC	vaccine. The polypeptides are encoded by NOX polynucleotides comprising	
CC	any of the 303 fully defined nucleotide sequences given in the	
CC	specification. The polypeptide is useful in the manufacture of a	
CC	medicament for treating a syndrome associated with a human disease. The	

```
Db 301 AAGAGGGGCTATACCTGCAAGGCTCCAGCGGATCACTGATGAAAGTGAGCAACTG 360
Qy 121 PheAlaAlATrPserAspHisArgProValCyArgAlaArgMetCysAspAlaHisIleu 140
Db 361 TTGCGCGGCTGGAGGCAACAGAGGCAAGTCTGGCCAGCCGCAATGTGATGCCAAGCTT 420
Qy 144 ArgGlyProSerGlyIleIleThrSerProanPheProIleGlnTrpAspAsnAla 160
Db 421 CGAGGCCCTCGGACATCACTCCCAATTTCCCAATTCCTCAAGTATGACAACATATCA 480
Qy 161 HisCysValTrpIleIleThrAlaLeuAnProSerIleValIleLeuAlaPheGlu 180
Db 481 CACTGTGTGTGATCATCACACACTCAACCCCTCCAGAGTATCAACTGCTTGAG 540
Qy 181 GluPheAspLeuGluArgGlyTrpAspThrLeuThrValGlyAspGlyGlnAspGly 200
Db 541 GAGTTGATTTGAGAGGGGCTATGACCCCTGAGCGTGGTGTGTGTGTGTGTGTGTGTGTGT 600
Qy 201 AspGlnLeuThrValLeuThrIleSerGlnAsnAlaCysSerAspSerProHisThrPro 220
Db 601 GACCAAGAGACAGCTTCTACATCTCTCAAAATGCTGACAGAGCCCTCACACACCA 660
Qy 221 GlySerArgIleProGluSerMetSerGlyAspIleTrpArgGlnLeuTrpThrValLeu 240
Db 661 GGCTCTCCATCCAGAGAGCATGTCTGGGAGCATCTGAGGAGAAATGACCTGTACTT 720
Qy 241 GluIleCysArgAspIleSerSerSerAspAlaArgSerGlySerValAlaGlySerPro 260
Db 721 GAGATCTCTGTGACATTAAGCATGATGATCAAGATCAAGATCAAGATCAAGATCAAGAT 780
Qy 261 LysThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlnGlnGlySerCys 280
Db 781 AAGCTTCTATAGCTGTGAGACTGTGTGCTCTGGAGACAGAGATCGAGAGGCAAGTTC 840
Qy 281 GlyAspProGlyIleProAlaTrpGlyAlaArgArgGluGlySerArgPheHisIleGlyAsp 300
Db 841 GGTGACCTGGCATACCTGATGATGCGCGAGGAGAGGATCCCGGTTTCCACCGGTGAC 900
Qy 301 ThrLeuIlePheGluCysGlnProAlaPheGluLeuValGlyGlnLeuAlaIleThrCys 320
Db 901 AACCTCAAGTTGAGTGCAGCGCCGCTTGGAGTGTGGAGACAGAGGCAATCACATGCG 960
Qy 321 GlnIleAsnAsnGlnTrpSerAlaLeuLeuProGlyCysValPheSerCysPhePheAsn 340
Db 961 CAAAGATATACCAATGCTGGCTAAGAGACAGGCTGCTGTCTCTGCTCTCTCTCTCTCAAC 1020
Qy 341 PheThrSerProSerGlyValValIleuSerProAnThrProGluAspTrpGlyAsnHis 360
Db 1021 TTTCACAGCCGCTGTGGGTTGTCTGTCTCTCCAACTACCAAGAGATATGCAACAC 1080
Qy 361 LeuHisCysValTrpLeuIleLeuAlaArgProGluSerArgIleHisIleuAlaPheAsn 380
Db 1081 CTCACAGTGTCTGTGCTCATCTGGCCAGGCTGAGGCGCATTCACCTGCGCTTCAAC 1140
Qy 381 AspIleAspValGluProGlnPheAspPheLeuValIleLeuAspGlyAlaThrAlaGlu 400
Db 1141 GACATTCAGCTGAGACCTCAAGTTGATTTCTGTGATCAAGATGAGGCGCACCCCGAG 1200
Qy 401 AlaProValLeuGlyThrPheSerGlyAsnGlnLeuProSerSerIleThrSerSerGly 420
Db 1201 GCCCGCGTCTGTGGGCACTTCTCAGGAAACAGCTTCCCTCTCTCACTCAAGCAGTGGC 1260
Qy 421 HisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLeuArgGlyPheAsnIle 440
Db 1261 CAGGTGACCGGCTTCAGATTCCAGACTGACACTCCACAGGAGAAAGGGGCTTCAACATC 1320
Qy 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyLeu 460
Db 1321 ACTTTTACACACTTCGACACAAAGAGTCCCGGATCTGGGCTTCCGTAATATGGCAA 1380
Qy 461 ArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGlyGlyPhe 480
1381 CGGTTTGGGAGACAGCCTCCAGCTGGGAGAGTCCATCTCTCTCTGTGATGAAGCTTC 1440
Qy 481 LeuGlyThrGlnGlySerGlyLeuThrIleThrCysValLeuLeuGlySerValValTrp 500
Db 1441 CTGGAGCTTCAGGGGCTCAAGAGACATCACCTGTGCTGATGAAGAGAGGAGCGGTGTGG 1500
Qy 501 AsnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisIleuThrSerProSerGly 520
Db 1501 AACAGGCTGTGTGTGGGTGTGAGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
Qy 521 ThrIleLeuSerProGlyTrpProGlyPheTrpLeuAspAlaLeuSerCysAlaTrpVal 540
Db 1561 ACCATCTCTCTCCGGGCTGGGCTGTGCTTCAAGATGCTTGTGAGTGTGTGTGTGTGTGT 1620
Qy 541 IleGluAlaGlyProGlyTrpProGlyLeuValIleThrPheAspArgPheLeuThrGluVal 560
Db 1621 ATTAGAGCCCAAGAGCTTACCCATCAAAATACCTTCGACATTCGAAATCAAAACAGAGTC 1680
Qy 561 AsnTrpAspThrLeuGluValAlaArgAspGlyArgThrTrpSerAlaProLeuIleGlyVal 580
Db 1681 AACTATGACACCTGGAAGTACGCGATGGGGGACTTACTACAGCCCTTGATCGGGGT 1740
Qy 581 TyrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAsnTrpLeuThrLeu 600
Db 1741 TACACGGGAGCCAGAGTTCCTCCAGTCTCTCATACAGACACACACATCACTCACTCTCTC 1800
Qy 601 PheSerThrAspLeuSerHisSerAspIleGlyPheGlnLeuArgTrpGluThrIleThr 620
Db 1801 TTCTTACCGACAAAGATCATCGGACATCGGCTTCCAGCTCCGCTATGAGACTATTAACA 1860
Qy 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHisIleGlyAsn 640
Db 1861 CTGCAGTACAGCACTGTCTGATCCAGAAATCCAGTAATGACAGCGTCAAGGGAAT 1920
Qy 641 AspPheTrpValGlyAlaLeuValThrPheSerCysAspSerGlyTrpThrLeuSerAsp 660
Db 1921 GACTTCTATCGTGGGCGGCTGTGATCTTCAAGCTGTGCTCGGCTCAACATTAAGTAC 1980
Qy 661 GlyGluProLeuGluCysGlnProAnPheGlnTrpSerArgAlaLeuProSerCysGlu 680
Db 1981 GGGAGCCTCTGGAGTGTGAGCCCACTTCAGTGGAGCGGGCGCTGCCAAGTTGTGA 2040
Qy 681 AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
Db 2041 GCTCTGTGTGTGCTTCAATCAAGGCTCCAGTGGGACATCTTGTGTCCAGAGGTTCTCT 2100
Qy 701 AspPheTrpProAnAsnLeuAsnCysThrTrpIleIleGluThrSerHisIleGlyLeu 720
Db 2101 GACTTCTAACCCCAACAATGAACTGAACTGCACTGGAATTAACAAATCTTCAATGCAAGGT 2160
Qy 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTrpLeuLeuIleThr 740
Db 2161 GTGTCTTCACTTCCACACCTTCCACTGGAAGAGGCGCATCACTCTCATCACT 2220
Qy 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
Db 2221 GAGAGGCGAGCTTCAACCCAGCCCTGAGGCACTTAAGTATGAGGCTGCGCAGCTCCC 2280
Qy 761 IleSerAlaGlyLeuTrpGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
Db 2281 ATACGCGCTGTGCTATGCACTTCACTGACCAAGTCCGCTTCACTCTGATTTCTCC 2340
Qy 781 MetSerTrpGluGlyPheAsnIleThrPheSerGluTrpAspLeuGluProCysGlyGlu 800
Db 2341 ATGTCTATTAAGATTAATCAATCACTTCTCAAGATGACACTTGGAGCGCTGTGGAG 2400
Qy 801 ProGluValProAlaTrpSerIleArgGlyLeuGlnPheGlyValGlyAspThrLeu 820
Db 2401 CCGAGGCTCCAGACTTACAGCATCCGGAAGGCTTGCACTTGGCTGGGCGACACCTTG 2460
Qy 821 ThrPheSerCysPheProGlyTrpArgLeuGlnGlyThrAlaArgIleThrCysLeuGly 840
Db 2461 ACCTTCTCTGCTTCCCGGGATCCGTCTGAGGGGACCGCGCATCACTGCTGGG 2520
```

QY 841 GIVARARGAArgLeuThrSerSerProleuProArgCysValAlaGluCysGlyAsnSer 660
 DB 2521 GGCGAGCGGGCGCTGTGAGAGCTCGCTCTGCCAAGGTGTGTGCTGAGATGGGGAATTC 2580
 QY 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnThrAsnAsn 880
 DB 2581 GTCAAGGCACTCAAGGTAATTTGTGTCTCCCAACTTCTCTGTGAATACATATACAT 2640
 QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysAlaArg 900
 DB 2641 CATAAATGCATCTCATCTCATCCAGACCAGGAGGAAAGGAAATTCACCTAAAGCCAGG 2700
 QY 901 AlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsnAsnSerLeu 920
 DB 2701 GCATTCGAAGCTCTCCGAAGAGATGCTCAAGGTTATGATGGCAACAACTCCGCC 2760
 QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerThrSer 940
 DB 2761 CGTTTGGCGGAGTTTAAAGCAATTCAGATGATGGGGGTGACTTTGAACAGCAATCC 2820
 QY 941 SerSerLeuThrPheAspPheIleThrAspAlaGluAsnThrSerLysGlyPheGluLeu 960
 DB 2821 AGCAGTCTGTGGCTTGAATTCATCACTGATGCTGAACACACAGCAAGGCTTTGAACCTG 2880
 QY 961 HisPheSerSerPheGluLeuIleLysCysGlyAspProGlyThrProLysPheGlyTyr 980
 DB 2881 CACTTTTCAGCTTTGAATCAATCAATGATGAGACCCAGAGAACCCCAATTTGGCTAC 2940
 QY 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
 DB 2941 AAGGTTCATGATGAAGATCATTTTGCAGGAGCTCCGTCTCTTCACTGATGACCTTGA 3000
 QY 1001 TyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyLysArgThrThrAsp 1020
 DB 3001 TACAGCCTCGCGGGAGTGAAGAGAGCTGTGTCTGATGAGAGACCGGACCTGGAGAC 3060
 QY 1021 ArgProLeuProThrCysValAlaGluCysGlyValThrValArgGlyValLeuSerGly 1040
 DB 3061 CGGCTCTCGCCACCTGTGTGCGCCAGTGTGAGAGGACAGTGAAGAGAGTGTGGGA 3120
 QY 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleThrThr 1060
 DB 3121 CAGGTGCTGTCAACCGGATATCCAGCTCCCTATGAACCAATCTCAATGATCTGACCC 3180
 QY 1061 IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
 DB 3181 ATCGAAGAGAGCGCGCTGCACCATTTGGCTACACTTCTGTGTGACACAGAGAG 3240
 QY 1081 ValHisAspValLeuArgIleThrAspGlyProValGluSerGlyValLeuLeuLysGlu 1100
 DB 3241 GTTACAGCAGTGTGCGCATCTGGATGGCTGTGAGAGCGGGGTCTCTGTAAAGAG 3300
 QY 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValValLeuGln 1120
 DB 3301 CTGAGTGGCCGGGCGCTGCCAAGGACCTGCATAGCACTTCAACTCGGTGTCTGTGAG 3360
 QY 1121 PheSerThrAspPhePheThrSerLysGlnGlyPheAlaIleGlnPheSerValSerThr 1140
 DB 3361 TTCAGCACTGACTTCTTCAACAGCAGCAGAGGCTTTGCCATTTTCACTGATGCCACA 3420
 QY 1141 AlaHisSerCysAsnAspProGlyIleProGluHisGlySerArgSerGlyAspSerThr 1160
 DB 3421 GCAACGTCTCTCAATGACCTGGAGTCCCGAGATGGGAGTCCGAGTGGAGACAGTTGG 3480
 QY 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGlySerAla 1180
 DB 3481 GAAGCGGCGACTCCACAGTGTTCAGTGTGACCTTGCTACGGCTGCAGAGAGATGCA 3540
 QY 1181 GluIleSerCysValLysIleGluAsnArgPhePheThrGlnProSerProProThrCys 1200
 DB 3541 GAGATCAAGCTGTGTGAAGATCGAAGAACAGTTCTTCTGTGAGCCAGCCGCCCAACATCG 3600

QY 1201 IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnThr 1220
 DB 3601 ATGCTCTCTCGGGGGAGACCTTACAGAACCAATCTGAGTCAATCTTCCACCAATTAC 3660
 QY 1221 ProGluProTyrProProGlyLysGluCysAspTrpLysValThrValSerProAspTyr 1240
 DB 3661 CCAGAACCTTACCCGCCAGGAGAGTGTGACTGGAAGTGAACGCTCTCACAGACTAC 3720
 QY 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
 DB 3721 GTCATTCGCCCTGGATTTAAATCATCTTAACTTGAGGCTGTGACTTCTCTCAATATC 3780
 QY 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGluLeuPro 1280
 DB 3781 TACAGCGAGCGGAGCTTCTTCAAGCTCTTCTTACAGAAAGCTTCTTGTGCTCCACTCCA 3840
 QY 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
 DB 3841 GGCCGCAATGAAACAGACAGACAGCAGCTTCTCTCGCTTCCGACAGATGATCTGTG 3900
 QY 1301 SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlyLysCysPheAspPro 1320
 DB 3901 AGCAATGCTGGCTTGTCAATGACTTATACAGAAACCCGCGGAGTCAATGTTTGAATCT 3960
 QY 1321 GlySerIleLysAsnGlyThrArgValGlySerAspLeuLysGlySerSerValThr 1340
 DB 3961 GGTTCATCAAGAACGGACAGCGGTGGGTCCACTGAAGCTGGGGCTCTCCGTCAAC 4020
 QY 1341 TyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysIleLeuGly 1360
 DB 4021 TACTACGCGCACCGGGGCTACGAGATGAGGACCTCGAACCTGAGCTGATCTGGGG 4080
 QY 1361 ProAspGlyLysProValTTPAsnAsnProArgProValCysThrAlaProCysGlyGly 1380
 DB 4081 CCTATGGGAAGCCCGGTGTGAACAAATCCCGGCAAGTGTGCACAGCCCCCTGTGGGGGA 4140
 QY 1381 GlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGluAsnTyrThrSer 1400
 DB 4141 CAGTATGGGTGTGAGAGAGTGTCTTGTCTCCCAATCAACCCCAAGATTAACCACT 4200
 QY 1401 GlyGlnIleCysLeuTyrPheValThrValProLysAspTyrValValPheGlyGlnPhe 1420
 DB 4201 GGACAGATCTGCTGTATTTTGTTCATCTGTGCCAAGGACTATGTGTGTCAGATTC 4260
 QY 1421 AlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGlnHis 1440
 DB 4261 GCCTTCTTTCACAGCGGCTTCAACAGAGTGTGAGGTTCCAGCGGCCACAGCACAGAC 4320
 QY 1441 SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeuAlaThr 1460
 DB 4321 TCGCGGCTCTCAAGCTCTCTCTCGGCTCCCATACAGAGATCACTGCTTGGCCACC 4380
 QY 1461 SerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHis 1480
 DB 4381 TCCATACAGTTCATTAAGTTCAGCGCCAAAGCTCTGCAGCAGCAGAGGCTTCCAC 4440
 QY 1481 PheValTyrGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
 DB 4441 TTTGTCTTACCAAGGGTTCCTCGAACCAGCGCCACGAGTGCAGCTGTGTGCCGAACC 4500
 QY 1501 ArgTyrGlyLysAspGluLysSerAspPheSerValGlyAlaIleValArgPheGluCys 1520
 DB 4501 CGTATGACAGAGAGCTGGGAGTGTCTCGGTGGGGGCACTGTGCTTGAATGC 4560
 QY 1521 AsnSerGlyTyrAlaLeuGlnGlySerProGluIleGluCysLeuProValProGlyAla 1540
 DB 4561 AACTCCGGCTATGCTCTGCAGGGGTCCCGCAGAGATCAAGTGTCTCTGTGTGCTGGGGCC 4620
 QY 1541 LeuAlaGlnTyrAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr 1560
 DB 4621 TTGGCCCAATGAAATGTCTCAGCGCCACAGTGTGTGTGCTGTGAGAGCAACTCACA 4680
 QY 1561 GluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrLeuAsnSerLeuAsn 1580


```

Db 4681 GAGCGCAGGGGACCATCTGTCCTCCGCTTCCAGAGCCCTTCAACAGCTCTCAAC 4740
Qy 1581 CyValATTPlyLeValValProGluGlyAlaGlyIleGlnIleGlnValValSerPhe 1600
Db 4741 TGGTGTGGAAGATCTGTGTCCCGAAGCGCTGGCATCAGATCCAGATTGTCTGATT 4800
Qy 1601 ValThrGluGlnAsnTrpAspSerLeuGluValPheAspGlyAlaAspAsnThrValThr 1620
Db 4801 GAGCAGAGCAGAACCTGGAGCTCGCTGAGAGTATTGATGGAGCAATTAACACTGTAACT 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGGAGTTTCTCAGAGAACACCGCTGCTGCTTCCGAAACAGACTCCAAACAG 4920
Qy 1641 LeuThrLeuHisPheThrSerAspIleSerValSerAlaAlaGlyPheHisLeuGluThr 1660
Db 4921 CTCTACCTTCAATTTCTACTCAGATATACAGCTATCTGACGTGGCTTCCACTTGGAGTAC 4980
Qy 1661 LysThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValIleThr 1680
Db 4981 AAAACGTGGGCTGAGCAGATTGTCTCGAACCTGCTGCTGCTGCTGCTGCTGCTGCTG 5040
Qy 1681 GlyGluArgTrpLeuValAsnAspValValSerPheGlnCysGluProGlyTrpValLeu 1700
Db 5041 GCGAGCCCTACTGTGTAATGATGTGTGTCTTCCAGTGTGAGCCGGAGATAGCCCTC 5100
Qy 1701 GlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTrpAsnTrpProPro 1720
Db 5101 CAGGGCCACGCCCAATCTCTGCAATGCCGGAACAGTGGGCGCATGAACTACCTCTCT 5160
Qy 1721 ProLeuCysIleAlaGlnCysGlyGlyThrValGluGluGluGluGlyValIleLeuSer 1740
Db 5161 CCACCTGTATATGACAGTGTGGGGGAAACAGTGGAGAGATGAGGGGGGTGATCTTGAGC 5220
Qy 1741 ProGlyPheProGlyAsnTrpProSerAsnMetAspCysSerTrpValIleAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCAACTACCCAGTAACATGAGACTGTCTCCGAAATAAGACACTGCC 5280
Qy 1761 ValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspTrp 1780
Db 5281 GTGGGCTTTGGAGCTCACTCACTCACTCTGTAACCTTCCACGAGCCCAACACGACTAC 5340
Qy 1781 IleGluIleArgAsnGlyProGlyTrpGluThrSerArgMetLeuValArgPheSerGlySer 1800
Db 5341 ATGAAATCCGGAATGGCCCTTATGAGACAGCCGATGATGGGAAGATTCTAGTGAAAC 5400
Qy 1801 GluLeuProSerSerLeuLeuSerThrSerHisGlyThrThrValTyrPheHisSerAsp 1820
Db 5401 GAGCTTCAAGCTCCCTCTCTCCAGCTGCCAGAACACCGTGTATTTCCACAGCCGAC 5460
Qy 1821 HisSerGlnAsnArgProGlyPheLeuGluThrGlnAlaTyrGluLeuGlnGluCys 1840
Db 5461 CACTCCCAAGATGGCCAGATTCAGAGCTGAGTATCAGGCCATGAACTTCAAGAGTGC 5520
Qy 1841 ProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrTrpAsnValGlyGln 1860
Db 5521 CAGACCCAGAGCCCTTGGCAATGGCATTTGTGAGGAGAGCTGGCTTACCACTGGAGACA 5580
Qy 1861 SerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys 1880
Db 5581 TCAGTACCTTCAGAGCTCCCGGGGTATCAATTGACTGAGCCACCTGCTCTCAAGTGT 5640
Qy 1881 GlnHisGlyThrAsnArgAsnTrpAspHisProLeuProValCysGluValProCysGly 1900
Db 5641 CAACAGGAGCAGCAACCGGAATGGAGCAGCCCTGCGCAAGTGGAGTCCCTTGTGTGC 5700
Qy 1901 GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTrpSer 1920
Db 5701 GGGAAATCATCTTCTTCCAGCAGCTGTACTCCCGGGGTCTCCCTTACGCCCTGACTCC 5760
Qy 1921 SerSerGlnAspCysValThrLeuIleThrValProIleGlyHisGlyValArgLeuAsn 1940

```

```

Db 5761 AGCTCCAGAGCTGTGTCTGCTGATCACCGTGCCTCATTTGGCATGGCGTCCGCTCAAC 5820
Qy 1941 LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln 1960
Db 5821 CTGAGCTGTGGAGCAGAGAGCCCTCTGGAGATTTTCATACCATCAGATGGGATGGCCAGAG 5880
Qy 1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaIleThrGlyGlyIlePheAla 1980
Db 5881 CAACAGCAGCAGAGGCTCGGCGCTTCAACCGGAGATGGCCAGAAAAACATGTCAGACT 5940
Qy 1981 SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaIleThrGlyGlyIlePheAla 2000
Db 5941 TCATCCACAGAGTCTCTGTCAAGTTCCACCGTATGAGCCACAGCGGGGGGATCTTGGCC 6000
Qy 2001 IleAlaPheSerAlaTyrProLeuThrLysCysProProProThrIleLeuProAsnAla 2020
Db 6001 ATAGCTTCTCGGCTTATCTACTACCAAAATCCCTCTCCACCATCTCCCAACGCC 6060
Qy 2021 GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTrpArgCysLeu 2040
Db 6061 GAAGTGTTCACAGAAATGAAGAAATTAATAGGTGACATCGTACGCTTACGATGCTC 6120
Qy 2041 ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrThrLeuGln 2060
Db 6121 CTGCTTCTTACTTGTGGGAAATGAATTTCTGACCTGCAAACTTGAACTTACCTGAG 6180
Qy 2061 PheGluGlyProProProIleCysGlyValHisCysProThrAsnGluLeuLeuThrAsp 2080
Db 6181 TTTGAAGACACACCCCGATATGTGAAGTGCACGTCCAAACAAATGAGCTTCTGACAGAC 6240
Qy 2081 SerThrGlyValIleLeuSerGlnSerTrpProGlySerTrpProGlnPheGlnThrCys 2100
Db 6241 TCCAGAGCGTGATCTTAGCAGAGCTACCTTGAGACTATCCCGAGTTCCACAGCTGC 6300
Qy 2101 SerTrpLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluThrPheLeu 2120
Db 6301 TCTTGCTGTGAGAGTGAGCCGAGCATATACATCTCCCTCACAGTGGAGTACTTCTC 6360
Qy 2121 SerGluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu 2140
Db 6361 AGCAGAGGAAGATGATGATTTGAGATTTTGATGTGCATCAGACAGAGCTCTCTG 6420
Qy 2141 LeuLysAlaLeuSerGlyAsnTrpSerAlaProLeuIleValThrSerSerSerAsnSer 2160
Db 6421 CTGAAAGCCCTCATGGGAATTACTAGCTCCCTGATTTGTCAACACTCAAGCAACTCT 6480
Qy 2161 ValTyrLeuArgTrpSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyr 2180
Db 6481 GTGTACCTGGTGTGTGATCTGATCAGCGCTTACAAATGGAAGGCTTCAAGATCCGCTAT 6540
Qy 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200
Db 6541 TCAGCCCTTACTGCACTGCGCCAGAGCTCACTCCATGAGCTTCTTCAAGCCAGAC 6600
Qy 2201 SerThrGlnProGlyLysSerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220
Db 6601 AGCACCAGCCCGGGGCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6660
Qy 2221 HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuThrSerGluAlaIle 2240
Db 6661 CACAGCATGGCATCTGTATCCCGGACCCCAAGGCTTACCACTGTGGAGCGCAAGCATTC 6720
Qy 2241 ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPhe 2260
Db 6721 CTCTCTGTCAACCTTTCTGTGTGGCTTCTGAGGCCCCCAAGATGGAATGGTGT 6780
Qy 2261 GlyLysGluThrThrValGlyThrLysAlaValTyrSerCysSerGlyGlyTyrHisLeu 2280
Db 6781 GGCAGAGAGTACACATGGAGAACCAAGGCCGTGTACAGCTGACATGGAAGCTTACCACTTC 6840
Qy 2281 GlnAlaGlyValGluAlaThrAlaGluCysLeuAspThrGlyLeuTrpSerAsnArgAsn 2300
Db 6841 CAGCAGGCGCTGAGGCCCTGCAAGTGTCTGACACAGGCGCTATGAGCAACCGCAAT 6900

```


QY 2301 ValProGlnCysValProValThrCysProAspValSerSerIleSerValGlnHis 2320
 DB 6901 GTCCACACACAGTGTGCTCCCTGTGACTGTCTGTATGTCAAGTACAGCTGAGAGAT 6960
 QY 2321 G1AgtTTPAgtLeuIlePheGlnThrGlnTyrglnPheGlnIleGlnLeuMetLeuIle 2340
 DB 6961 GGCCGATGAGGCTTATCTTTGAGACACAGTATCAGTTCAGAGCCACGCTATGCTCATC 7020
 QY 2341 CysAspProGlyTyrTyrTyrTyrThrGlnGlnArgValIleArgCysGlnIleAsnGlyLys 2360
 DB 7021 TGTGACCTGGCTACTACTATACGTGGCAAGGGTCATCCGCTGCAGGCCAATGGCAAA 7080
 QY 2361 TrpSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGlnLeuProIle 2380
 DB 7081 TGGAGCTCGGGGACTTACCCGCCACTGCCAATCATCTCTGTGGAGAGCTCCCATTT 7140
 QY 2381 ProProAsnGlyHisArgIleGlyThrLeuSerValTyrglyValAlaThrAlaIlePheSer 2400
 DB 7141 CCCCCAATGGCCACCGCATCGGAACATGTCGTCTAGCGGGGCAACAGCCATCTTCTCC 7200
 QY 2401 CysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu 2420
 DB 7201 TGCATTCGCGATACACACTGGTGGCTCCAGGGGCGGTGATGTCATGGCCAAATGGGCTC 7260
 QY 2421 TrpSerGlySerGluValArgCysLeuAlaGlyHisCysGlyTyrProGluProIleVal 2440
 DB 7261 TGGATGGCTCTGAAGTCCGCTGCTTGTGTGACACTGTGGAGCTCTGAGCCCATTTCTC 7320
 QY 2441 AsnGlyHisIleAsnGlyGlnAsnTySerTyArgGlySerValValTyrglnCysAsn 2460
 DB 7321 AACGGACATCAATGGGGAGAACATACAGCTACCGGGGCAGTGTGTTCACCAATGCAT 7380
 QY 2461 AlaGlyPheAlaGlyLeuIleGlyMetSerValArgIleCysGlnIleAsnHisIleTrpSer 2480
 DB 7381 GCTGGCTTCCGCTCATCGGCTGTCTGTGGCATCTGCCACAGAGATCATCAGTGGCTG 7440
 QY 2481 GlyLeuThrProPheCysValProIleThrCysGlyHisIleProGlyAsnProValAsnGly 2500
 DB 7441 GGCAAGACCCCTTTTGTGTGCAATTACTGTGGACACCAAGGCACCCCTGTCAACGGC 7500
 QY 2501 LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValCysAsnProGly 2520
 DB 7501 CTCACTACGAGGTAAACAGTTTAACCTCAACGATGTGTCAGTTGTTTGGCAACCTTGGG 7560
 QY 2521 TyrMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTrpSerAspMet 2540
 DB 7561 TATATGGCTGAGGGGCTGTAGTCCCAATGCCCTGGCCAGCGGCAATGAGTGCATG 7620
 QY 2541 LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnAsnSerValArg 2560
 DB 7621 CTGCCACCTGTGAGATATCATCACTGATCAAGATCTGTGACACCAAGAAATAGTGTGT 7680
 QY 2561 GlnValHisAlaSerGlyProHisArgPheSerPheGlyTyrThrValSerTyArgCys 2580
 DB 7681 CAGGTCCACGCGAGGGGCCGCGACAGTTCAGCTTCGGCACCACTGTGTCTTACCGGGTC 7740
 QY 2581 AsnHisGlyPheTyLeuLeuGlyTyrProValLeuSerCysGlnGlyAspGlyTyrTrp 2600
 DB 7741 AACCAAGGCTTCTACTCTCTGGGCAACCAGTGTCACTGCAAGGAGAGATGGCAATGG 7800
 QY 2601 AspArgProArgProGlnCysLeuLeuValSerCysGlyHisIleProGlySerProProHis 2620
 DB 7801 GACCTTCCCCCGCCAGATGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 7860
 QY 2621 SerGlnMetSerGlyAspSerTyThrValGlyAlaValAlaArgTySerCysIleGly 2640
 DB 7861 TCCCAAGATGTCTGAGACAGTATATCTGTGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 7920
 QY 2641 LysArgThrLeuValGlyAsnSerThrArgMetCysGlyLeuAspGlyHisIleTrpThrGly 2660
 DB 7921 AAGCTACTCTGTGGGAAACAGACCCGCACTGTGTGGCTGTGATGAGACATGAGCTGGC 7980

QY 2661 SerLeuProHisCysSerGlyTyrThrSerValGlyValCysGlyAspProGlyIleProAla 2680
 DB 7981 TCCCTCCCTCACTGCTCAGGACCAAGCTGGAGATTGGCGTGAACCTGGGATCCCGCT 8040
 QY 2681 HisGlyIleArgLeuGlyAspSerPheAspProGlyTyrValIleArgPheSerCysGlu 2700
 DB 8041 CATGGCATCCGTTTGGGGGAGAGCTTGTGATCCAGGCACTGTGATGCCCTTCAGCTGTGAA 8100
 QY 2701 AlaGlyHisValLeuArgGlySerSerGlyAspThrThrCysGlnIleAsnGlySerTrpSer 2720
 DB 8101 GCTGGCCACGCTCCCGGGATGCTCAGAGGCACTGTCAAGCCAAATGGCTGTGGAGC 8160
 QY 2721 GlySerGlnProGluCysGlyValIleSerCysGlyAsnProGlyTyrProSerAsnAla 2740
 DB 8161 GGCTCGAGCTTAGTGTGAGATGATCTCTTGTGGAAACCTGGGACTCCAAATGAATGCC 8220
 QY 2741 ArgValAlaPheSerAspGlyLeuValPheSerSerIleValTyrgluCysArgGlu 2760
 DB 8221 CGAGTTGTGTAGTATGAGCTGTGTTTCTCCAGCTCTATGCTATGATGTCGGGAA 8280
 QY 2761 GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyTyrTrpThrGly 2780
 DB 8281 GGATATACCGCACAGGCTGTCTACGCTACGTGTGGTCAATGGTATCTGAGACAGC 8340
 QY 2781 SerAspProGluCysLeuValIleAsnGlyAspProGlyIleProAlaAsnGlyLeu 2800
 DB 8341 AGTACCTTAGTCCCTCGTCAATTAACGTGTGTGACCTGGAGATCCAGCCAAATGGCTT 8400
 QY 2801 ArgLeuGlyAsnAspPheArgTyAsnLysThrValThrTyrglnCysValProGlyTyr 2820
 DB 8401 CGGCTGGGCAATGACTTCAGGTACAAACAAATGTGTACATATCAGTGTGCCCTGCTAT 8460
 QY 2821 MetMetGluSerHisArgValSerValLeuSerCysThrLysAspArgThrTrpAsnGly 2840
 DB 8461 ATGATGAGATCAATAGATATCTGTGTGACCTGACCAAGAGCCGGAATGGAATGGA 8500
 QY 2841 ThrLysProValCysLysAlaLeuMetCysLysProProLeuIleProAsnGlyLys 2860
 DB 8501 ACCAAGCCGCTGTGCAAGCTCTCATGTGCAAGCACTCCGCTCAATCCCAATGGGAA 8560
 QY 2861 ValValGlySerAspPheMetTrpGlySerSerValThrTyralaCysLeuGluGlyTyr 2880
 DB 8561 GTGGTGGGATCTGACTCATGTGGGCTCAAGTGTGACTTATGCTCTCGAGGGGTAC 8640
 QY 2881 GlnLeuSerLeuProAlaValPheThrCysGluGlyAsnGlySerTrpThrGlyGlnLeu 2900
 DB 8641 CAGCTTCTCTGCCCCGGGTGTTCACCTGTGAGGAATAGGTCTGAGCCGAGAGCTG 8700
 QY 2901 ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgGlyArgGlu 2920
 DB 8701 CTTCAAGTCTTCTCTGTGTCTGTGGGAAATCTGTGTCTCCGTCCTGGGAGAGAGAG 8760
 QY 2921 AspArgGlyPheSerTyArgSerSerValSerPheSerCysHisProProLeuValLeu 2940
 DB 8761 GACGAGAGCTTCTCTCAAGTATATCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8820
 QY 2941 ValGlySerProArgArgPheCysGlnSerAspGlyTyrTrpSerGlyTyrGlnProSer 2960
 DB 8821 GTGGGCTCTCCACCAAGTATTTGGCACTGAGTGGGCAATGAGGTGCAACCGCCAGC 8880
 QY 2961 CysIleAspProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyIleGln 2980
 DB 8881 TGCATTAATCCAGCCCTGACCAAGTGTGGGACCTGTGTGTGCAAGTTTGGATACAG 8940
 QY 2981 AsnAsnSerGlnGlyTyrglnValGlySerThrValLeuPheArgCysGlnGlyTyr 3000
 DB 8941 AACATTCCTACGAGCTACAGGTTGGAAGCAAGTCTCTTCGGTTGTCAAAAAGGCTAC 9000
 QY 3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTrpSerGlyTyrPro 3020
 DB 9001 CTGCTTCAAGGCTTCCACACAGGACCTGTGCTCCAAACCTGAGTGTGAAACCCCA 9060
 QY 3021 ProAspCysValProHisHisCysArgGlnProGluThrProThrHisAlaAsnValGly 3040

Db 9061 CCTGACGTGTCCCCCACTGCAGCAGCAGACGCCAAGCATGCAAGTCGGG 9120
Qy 3041 AlaleuspleuProsaMetGlyThrThreuleTherProAlaagaAlaSerPro 3060
Db 9121 GCCCTGATTGCTCTCCACAGGGCTACAGCTCATTTCTCTGCAAGAGGCTTCTCC 9180
Qy 3061 SerArgValAlaProSerThrAlaProAlaArgMetAlaAlaGlyGlnAlaSerArg 3080
Db 9181 TCAAGGCTGCTCCGACACCGACCTGCAAGGCGATGCGAGCTGACAGGCAAGCCGC 9240
Qy 3081 ProSerAlaTrpArgSerGlyProValGlyAspProSerThrleuProGlySerHArg 3100
Db 9241 CCACTGCTCGAGAGTCCGGCCCACTGGAGAGACCATCAACTGCCCCGGAGCCACCGC 9300
Qy 3101 SerProlyPro 3104
Db 9301 TCACCCAGGCT 9312

RESULT 3
ADH71137
ID ADH71137 standard; DNA; 10136 BP.
XX AC ADH71137;
XX 25-MAR-2004 (first entry)
DE Human gene of the invention NOVab SEQ ID NO:33.
XX de; gene; human; cytostatic; immunomodulator; neuroprotective; nootropic;
XX anorectic; antidiabetic; antimicrobial; antihypertensive; gene therapy;
XX cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX obesity; diabetes; infectious disease; metabolic syndrome X;
XX dyslipidemia.
OS Homo sapiens.
XX NC02003102155-A2.
XX 11-DEC-2003.
PD 03-JUN-2003; 2003MO-US017430.
XX 03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 05-JUN-2002; 2002US-038647P.
PR 06-JUN-2002; 2002US-0386376P.
PR 06-JUN-2002; 2002US-0386453P.
PR 06-JUN-2002; 2002US-0386864P.
PR 06-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386816P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0386942P.
PR 07-JUN-2002; 2002US-0386971P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-0296960P.
PR 10-JUN-2002; 2002US-0387400P.
PR 10-JUN-2002; 2002US-0387535P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.

PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389848P.
PR 19-JUN-2002; 2002US-0390066P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 12-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 12-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0406317P.
PR 15-AUG-2002; 2002US-0406317P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.
XX Alsbrogk JP, Alvarez E, Anderson DM, Boldog FL, Casman SJ;
PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
PI Eitenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
PI Gusev VI, Herrmann UL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
PI MacIsachlan T, Malyskar M, Mezick AJ, Millet I, Mishra VS;
PI Padigaru M, Patuturajan M, Pena CE, Peyman JA, Raha D, Rastelli L;
PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA;
PI Smithson G, Spylek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M;
PI Zhong H;
XX WPI, 2004-081935/08.
DR P-PSDB; ADH71138.
XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX Example 4; SEQ ID NO 33; 1880bp; English.
XX The invention relates to a novel isolated polypeptide (NOVX). A
CC polypeptide of the invention has cytostatic, immunomodulator,
CC neuroprotective, nootropic, antidiabetic, antimicrobial, and
CC anti-inflammatory activity, and may have a use in gene therapy, and as a
CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising

any of the 303 fully defined nucleotide sequences given in the specification. The polypeptide is useful in the manufacture of a medicament for treating a syndrome associated with a human disease. The polypeptide, polynucleotide and antibody are useful in diagnosing, treating or preventing NOX-associated disorders, e.g. cancer, cachexia Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are further used as hybridisation probes, in chromosome mapping, tissue typing, preventative medicine, and pharmacogenomics. The present sequence encodes a NOX polypeptide of the invention.

Sequence 10136 BP; 2147 A; 3060 C; 2744 G; 2185 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	10136
Score:	16985.00	Matches:	31040
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	12	Gaps:	0

US-10-016-248-2 (1-3104) X ADH71137 (1-10136)

[illegible]

QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyValArgHisGlyAsn 640
 DB 1861 CTGCACATCAGACCACTGCTCGATCCAGAAATCCAGTAAATGACACCGTCATAGGGAAT 1920
 QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660
 DB 1921 GACTTCTACGTGGGGCGCTGGTGAACCTTCAGCTGTGACTGGGGTACATTAAGTAC 1980
 QY 661 GlyGluProLeuGluCysGluProAsnPheGlnTyrPheArgAlaLeuProSerCysGlu 680
 DB 1981 GGGAGACCTCTGAGTGAAGCCCAACTTCAGTGAAGCCGGGCCCTCCAGTGTGGAA 2040
 QY 681 AlaLeuCysGlyGlyPheIleGlnIlySerSerGlyThrIleLeuSerProGlyPhePro 700
 DB 2041 GCTCTGTGGTGGCTTCATTCAGAGCTCCAGTGGACCATCTTGCCAGGGTTCCT 2100
 QY 701 AspPheTyrProAsnAsnLeuAsnCysThrTyrIleIleGluThrSerHisGlyLysGly 720
 DB 2101 GACTTCTACCCCAACACCTTGAACCTGACCTGATATCGAAACATCTCATGGCAAGGT 2160
 QY 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyrLeuLeuIleThr 740
 DB 2161 GTGTTCTTCACCTTCCACACCTTCACCTGGAAAGTGGCCATGACTACCTTCATCACT 2220
 QY 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
 DB 2221 GAGAACGGACCTTCACCCACGACCCCTGAGGAGCACTGATCTGGGCTGCCAGCTCCC 2280
 QY 761 IleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
 DB 2281 ATCAGCGCTGGGCTCTTATGGCACTTCACTGCCCAAGGTCCGCTTCATCTGATTTCTCC 2340
 QY 781 MetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCysGluGln 800
 DB 2341 ATGTCATATGAAAGATTCACATCACTTCTCAGAGTACACATCTTGAGCCCTGTAGAG 2400
 QY 801 ProGluValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeu 820
 DB 2401 CCCAGAGTCCAGCTTCACACATCCGGAAGGGCTTGCACTTGCGGGCGACCACTTG 2460
 QY 821 ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840
 DB 2461 ACCTTCCTCGCTTCCCGGGGATCCGCTGAGAGGACCGCCGCAATCACTGCTGGGG 2520
 QY 841 GlyArgArgArgLeuTyrPheSerProLeuProArgCysValAlaGluCysGlyAsnSer 860
 DB 2521 GGCAGAGGGGCGCTGTGAGACTCGCTGCAAGGTGTGTGCTGAGTGGGAATTC 2580
 QY 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn 880
 DB 2581 GTCAACAGGCACTCAGGGTACTTGTGCTGCCCAACTTCTGTGAACTCAATTAACAT 2640
 QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysAlaArg 900
 DB 2641 CATGAATGATCTACTACATCCAGACCCAGCCAGGAAAGGAATTCAGTGAACCCAG 2700
 QY 901 AlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsnAsnAsnSerAla 920
 DB 2701 GCAATTGAACCTCCCAAGAGATGTCTCAAGGTTTATGATGGCAACAACAATCCGCC 2760
 QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerThrSer 940
 DB 2761 CGTTGCTGGAGTTTATGCCATTCGTGAGTGAATGGGGGTACTTTGAACAGCAATCC 2820
 QY 941 SerSerLeuTyrPheAspPheIleThrAspAlaGluAsnThrSerGlyPheGluLeu 960
 DB 2821 AACAGCTGTGGCTGATTCATCACTGATCTGAAAAACACAGCAAGGGCTTTGAATG 2880
 QY 961 HisPheSerSerPheGluLeuLeuLysCysGlyAspProGlyThrProLysPheGlyTyr 980
 DB 2881 CACTTTTCCAGCTTGAACCTCAATAATGTGAGACCCAGAAACCCCAAGTTTGGCTAC 2940

QY 981 LysValHisAspGlyLysIlePheAlaGlySerSerValSerPheSerCysAspProGly 1000
 DB 2941 AAGTTCATGATGAAGTCAATTTTGCAGGAGACTCCGTCTCTTCACTGATCACTTGA 3000
 QY 1001 TyrSerLeuArgGlySerGluLeuLeuCysLeuSerGlyGluArgArgThrPheAsp 1020
 DB 3001 TACAGCTCTGGGGTGAAGTGAAGAGCTGTGTCTGATGGTGAAGACCGCCGACTGGAG 3060
 QY 1021 ArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluValSerGly 1040
 DB 3061 CGGCTCTGCCACCTGTCTCCCGAGTGTGAAGGACATGAGAGAGAGGTGTCCGGGG 3120
 QY 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleThrThr 1060
 DB 3121 CAGGTGCTGTCAACCGGGATATCACTCTCTGAAACAATCTCACTGATCTGGACC 3180
 QY 1061 IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
 DB 3181 ATCGAAGCAGAGCGCGCTGACCACTTGGCTTACCTTCTGTGTGTGACACAGAGAG 3240
 QY 1081 ValHisAspValLeuArgIleThrAspGlyProValGluSerGlyValLeuLeuLysGlu 1100
 DB 3241 GTTCACAGAGTCTGGCATCTGGGATGGGCTGTGGAGAGCGGGTCTTCTGAAGAG 3300
 QY 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValIleGln 1120
 DB 3301 CTGAGTGGCGGGCTCTCCCAAGACCTGTATGACACTTCACTCGGTGCTCGAG 3360
 QY 1121 PheSerThrAspPhePheThrSerLysGlnIlyPheAlaIleGlnPheSerValSerThr 1140
 DB 3361 TTCAGCACTGACTTCTTACACAGACAGAGGGCTTTGCCATTTCAATTTTCAAGTCCACA 3420
 QY 1141 AlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSerGlyAspSerTyr 1160
 DB 3421 GCACGTCTTCAATGACCTTGGGATCCCGCAAGATGGAGTGGAGTGGTGAACATTTGG 3480
 QY 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGlySerAla 1180
 DB 3481 GAAGCCGGCGACTCACAGTGTTCACATGTGAACCTTGGCTGACGGGAGGAAAGTGA 3540
 QY 1181 GluIleSerCysValLysIleGluAsnArgPhePheThrGlnProSerProProThrCys 1200
 DB 3541 GAGATCAGCTGTGGAATGAGAAACAGGTTCTTGGACGCCAGCCGCCAACATGC 3600
 QY 1201 IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
 DB 3601 ATCCCTCCCTGGGGGAGACTGACAGACCATCTGGAGTCACTCTTCAACCAAAATTAC 3660
 QY 1221 ProGluProTyrProProGlyLysGluCysAspTyrPheValThrValSerProAspTyr 1240
 DB 3661 CCAGAACCTTACCCGCCAGCAAGAGTGTGACTGAAAGTGCCTGTTCACCAAGACTAC 3720
 QY 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
 DB 3721 GTATTCGCCCTGTGATTTAACTTTAACTTGAACCTGGAGCTGTGATGACTTCTCCATATC 3780
 QY 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro 1280
 DB 3781 TACGACGAGAGGAGCTCTTCAACCTCTCATAGGAAGCTTCTTATAGGCTCCAGCTCCA 3840
 QY 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
 DB 3841 GGCCTGATGAAAGCAGACCAACAGCTCTTCTGCTTCCGACGATGATCACTGTG 3900
 QY 1301 SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPheAspPro 1320
 DB 3901 AGCAATGCTGGCTTCTCATTTACTATACAAAAACCCCGGAGATGATGTTTATCTCT 3960
 QY 1321 GlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSerValThr 1340
 DB 3961 GGTTTCATCAAGACGACACAGGGGTGGGTCCGACTGAACTGGCTCTCCCTGCATCC 4020
 QY 1341 TyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysIleLeuGly 1360

```
Db 4021 TACTACTGCGACGGGGGCTGACGAGTTGAGGGGACCTGACCTGAGCTGCATCTGGG 4080
Qy 1361 ProAspGlyLeuProValITPAsnAsnProArgProValGlyThrAlaProCysGlyGly 1380
Db 4081 CCTGATGGGAAGCCCGTGTGAAACAATCCCGGCGACGTGACACGCCCTGTGGGGAGA 4140
Qy 1381 GlnTyValGlySerAspGlyValValLeuSerProAsnTyProGlnAsnTyThrSer 1400
Db 4141 CAGATGTGGGTTGCGACGAGATGATCTTGTGCCCACTACACCACTACCACTACCACT 4200
Qy 1401 GlyGlnIleCysLeuTyPheValThrValProIleAspTyValValPheGlyGlnPhe 1420
Db 4201 GGACGATCTGCTGTATTTTGTACTGTGCCAAGACTATGTGTGTGGCCAGTTC 4260
Qy 1421 AlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGlnHis 1440
Db 4261 GCCTTCCTTTCACACGCGCTTCAACGACGTGTGGAGTTCAACGCGCCACAGCCAGCAC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGlySerLeuProLeuAlaThr 1460
Db 4321 TCGGGGCTCTCAAGCTCCCTCTCGGGCTCCATACAGAGAAATCACTGCCCTTGGCCACC 4380
Qy 1461 SerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHis 1480
Db 4381 TCCATCAAGTTCTCATTAAGTTACGGCCAAAGGCTCGCACACAGCCAGAGGCTTCAC 4440
Qy 1481 PheValTyArgAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
Db 4441 TTTGTCTTCCAAAGCGGTTCTCTGAAACGCGCCAGCGACGTGCTGTGCTCGGAAACC 4500
Qy 1501 ArgTyArgLysArgLeuGlySerAspPheSerValGlyAlaIleValArgPheGluCys 1520
Db 4501 CGCTATGGCAAGAGCTGGGAGTACTTCTGGAGGGGCGATGTCGCTGCAATGC 4560
Qy 1521 AsnSerGlyTyAlaLeuGlnGlySerProGluIleGluCysLeuProValProGlyAla 1540
Db 4561 AACTCGGCTATGCCCTGACAGGGGTCCGACAGATCGATGCTCCCTGCTGGGGCC 4620
Qy 1541 LeuAlaGlnTPAsnValSerAlaProThrCysValValProCysGlyValAsnLeuThr 1560
Db 4621 TTGGCCCAATGAAATGTCTACGCGCCACGTGTGTGGTGGAGGAGCAACTCACA 4680
Qy 1561 GluArgArgGlyThrIleLeuSerProGlyPheProGluProIleLeuAsnSerLeuAsn 1580
Db 4681 GAGGCGAGGGGACCATCTGTCCTGCTGCTCCAGAGCGTACCTAAACAGCCTCAAC 4740
Qy 1581 CysValITPLeuValIleValValProGluGlyIleGlnIleGlnValIleValSerPhe 1600
Db 4741 TGTGTGTGGAAGATGTGTGTCGCGAAGGCGCTGGCATCCAGATCCAGTTGTCAATT 4800
Qy 1601 ValIThrGluGlnAsnTPAsnSerLeuGluValPheAspGlyAlaAsnAsnThrValThr 1620
Db 4801 GTGACAGAGCAAGACTGGGACTCGCTGGAAGATTTGATGTGTCAGATTAACACTGTAAAC 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGAGATTTCTCAGAAACAACGTCGCTGCTCTGAAACAGACCTCCAAACAG 4920
Qy 1641 LeuTyIleuHisPheTySerAspIleSerValSerAlaAlaGlyPheHisLeuGluTyI 1660
Db 4921 CTTCACCTTCAATTTCTACTCAGATATACGCGATATCGAGCTGGCTTCATTGGAGTAC 4980
Qy 1661 LysThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValIleThr 1680
Db 4981 AAAACGGTGGGCTGAGCAGATGTGTCGGAACCTGCTGTGCCAGTAAAGGGGTGAAGACT 5040
Qy 1681 GlyIleuArgTyIleuValAsnAspValValSerPheGlnCysGluProGlyTyAlaLeu 1700
Db 5041 GCGAGGCTTACTGTGTAATGATGTGTGTCTTTCAGTGTGAGCCGGGATATGCTCTC 5100
Qy 1701 GlnGlyHisAlaHisIleSerCysMetCProGlyThrValArgArgTPAsnTyProPro 1720
Db 5101 CAGGGCCACGCCCATCTCTCTGATGCCGGAACAGTGGCGCGATGAACTACCTCTCT 5160
Qy 1721 ProLeuCysIleAlaGlnCysGlyGlyThrValGluGluMetGluGlyValIleLeuSer 1740
Db 5161 CCACCTGTATATGACAGTGTGGGGAACAGTGAAGAGATGAGGGGGTGTATCTGAC 5220
Qy 1741 ProGlyPheProGlyAsnTyProSerAsnMetAspCysSerITPLeuValIleAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCACTACCCAGTAACTGAGTCTCTCTGAAAAATAGCATCTGCC 5280
Qy 1761 ValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspTyI 1780
Db 5281 GTGGGCTTGGAGCTACATCTCAGTTCTCTCAACTTCTCCACCGAGCCCAACAGCATAC 5340
Qy 1781 IleGluIleArgAsnGlyProTyGluIThrSerArgMetMetGlyArgPheSerGlySer 1800
Db 5341 ATGAAATCCGGAATGGCCCTTAAGAACACGCGCATGATGGGAATTCAGTGAAGC 5400
Qy 1801 GluLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTyPheHisSerAsp 1820
Db 5401 GAGCTTCCAAAGCTCCCTCTCTCCACGTCACAGACACACCGTGTATTTCCACAGCGAC 5460
Qy 1821 HisSerGlnAsnArgProGlyPheLysLeuGluTyGlnAlaTyGluLeuGlnGlyCys 1840
Db 5461 CACTCCAGAAATCGGCGAGATTCAGCTGGAATACAGGCTGTGAACCTTCAAGAGTGC 5520
Qy 1841 ProAspProGluProPheAlaAsnGlyIleValArgGlyValArgTyIleAsnValGlyGln 1860
Db 5521 CCAGACCCAGAGCCCTTGGCAATGGCATGTGGAGGAGACTGGCTTCAACGTGGGCAAA 5580
Qy 1861 SerValIThrPheGluCysLeuProGlyTyGlnLeuThrGlyHisProValLeuThrCys 1880
Db 5581 TCACTGACCTTCAGATGCTCTCCGGGATATCAATGACTGGCCACCTGTCTCACTGTGT 5640
Qy 1881 GlnHisGlyIThrAsnArgAsnTPAsnAspProLeuProLysCysGluValProCysGly 1900
Db 5641 CAACATGACACCAACCGGAATGGGACCAACCCCTGCCCAAGTGTGAAGTCCCTTGGGC 5700
Qy 1901 GlyAsnIleThrSerSerAsnGlyThrValTySerProGlyPheProSerProTySer 1920
Db 5701 GGGAAACATCACTTCTTCCAAAGGACGTGTATCTCCCGGGGTTCCTTACCCCTACTCC 5760
Qy 1921 SerSerGlnAspCysValITPLeuIleThrValProIleGlyHisGlyValArgLeuAsn 1940
Db 5761 AGCTCCAGAGACTGTGTGCTGATATCAACGTGCCCATTTGGCATGGCGCTCAAC 5820
Qy 1941 LeuSerLeuLeuGlnIThrGluProSerGlyAspPheIleThrIleTPAspGlyProGln 1960
Db 5821 CTGAGCTGTGAGACAGAGCCCTGTGAGATTTATCAACATCTGGATGGGCCACAG 5880
Qy 1961 GlnIThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLysIThrValGlnSer 1980
Db 5881 CAACACACACACCGGCTCGGCTCTTCAACCGGAGCTGGCCAAAGAAACAGTGCAGAGT 5940
Qy 1981 SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaIleAlaIThrGlyGlyIlePheAla 2000
Db 5941 TCATCAACAGATCTCTGCTCAAGTTCCACCGTGATCAAGCCACAGAGGGGAGTTTGGC 6000
Qy 2001 IleAlaPheSerAlaTyProLeuIThrLysCysAspProProIThrIleLeuProAsnAla 2020
Db 6001 ATAGCTTCTTCGCTTATTCACCTACCAAAATGCCCTCTCCACACATCTCCCAAGGCC 6060
Qy 2021 GluValValIThrGluAsnGluGluPheAsnIleGlyAspIleValAlaGlyTyArgCysLeu 2040
Db 6061 GAAGTCTGTACAGAGATGAAGAAATTCAAATATGTGTACATCGAACCTTACAGATGCTC 6120
Qy 2041 ProGlyPheIThrLeuValGlyAsnGluIleLeuThrCysLysLysGluIThrTyLeuGln 2060
Db 6121 CTGGCTTTTACCTTAGTGGGGAATGAATTCAGACCTGCAAACTTGGAACTTACTCTCAG 6180
Qy 2061 PheGluGlyProProIleCysGlyValHisCysAspProIThrAsnGluLeuLeuThrAsp 2080
Db 6181 TTGGAAGACCAACCCGATATGTGAAGTGCATGTCCAAATATGAGCTTCTGACAGAC 6240
```

```
QY 2081 SerThrGlyValIleuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys 2100
DB 6241 TCACAGGGCTGATCTCTGAGCCAGAGCTACCTGGAGAGCTATCCCAAGTTCACAGCCGCG 6300
QY 2101 SerTyrPheValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu 2120
DB 6301 TCTTGCGCTGGTGAGAGTGGAGCCGACATATACATCTCCCTCAGAGTGGATATCTTCTCC 6360
QY 2121 SerGluValGlnTyrAspGluPheGlnIlePheAspGlyProSerGlyGlnSerProLeu 2140
DB 6361 ACCGGAAGACATATAGATGAGTTTGAATTTTGAATGCTCATAGAGACAGAGCTCTG 6420
QY 2141 LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer 2160
DB 6421 CTGAAGCCCTCAGTGGGAAATTACACACTCCCTGATTCACAGCTCAAGCAACTCT 6480
QY 2161 ValTyrLeuAlaGlyTyrPheSerThrAspHisAlaTyrAsnArgGlyPheValIleArgTyr 2180
DB 6481 GTGTAACCTGGCTGGTCTCATCTGATCAGCCCTACATCGAAGGGCTTCAGATCCGCTAT 6540
QY 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200
DB 6541 TCAGCCCTTACTGACAGCTCTGACAGGAGCTCCACTCCATGAGCTTCATCTAAGGCCAGACC 6600
QY 2201 SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220
DB 6601 AGCACCCAGCCGGGGGCTCCATCCACTTGGCTGCAACGCGGCTACCCCTGGTGGGA 6660
QY 2221 HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrPheSerGluAlaIle 2240
DB 6661 CACAGCATGGCCATCTGTACCCGGGACCCCGGAGGCTACCACTGTGGAGCCGAAGCCATC 6720
QY 2241 ProLeuGlnAlaLeuSerCysGlyLeuProGlnAlaProLysAsnGlyMetValPhe 2260
DB 6721 CCTCTCTGTCAAGCTCTTCTCTGTGGCTTCTGTAGGCCCCCAAGATGGAAAGTGATTT 6780
QY 2261 GlyValGluTyrThrValGlyTyrLysAlaValTyrSerCysSerGlyGlyTyrHisLeu 2280
DB 6781 GGCAGAGAGTACACAGTGGGAAACCAAGCGCTGTACAGTGCAGTGAAGGCTACCACTC 6840
QY 2281 GlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuTyrPheAsnArgGln 2300
DB 6841 CAGGACAGGCTGTAGGACACTGTAGAGTGTGTGACACAGGCTTATGAGCAACCCCAAT 6900
QY 2301 ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGluHis 2320
DB 6901 GTCCACACACAGTGTGTCTGTGAGCTTGTCTGATGTGATGATGATGATGATGATGATG 6960
QY 2321 GlyArgThrArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle 2340
DB 6961 GGCAGATGGAGGCTTATCTTTGAGACACAGTATCAGTTCAGGCCCCAGCTGATGCTCATC 7020
QY 2341 CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys 2360
DB 7021 TGTGACCTCGGCTACTATCTACTGTCGCAAGGGTATCATCGCTGTGAGCCATGGCAAA 7080
QY 2361 TyrSerLeuGlnLysSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle 2380
DB 7081 TGGAGGCTCGGGGACCTTACGCCACCTGCCGATCATCTCTGTGGAAGGCTCCGAGTT 7140
QY 2381 ProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSer 2400
DB 7141 CCCCCCAATGGCCACGCGATCGGAACATGCTGTGCTTACGGGGCAACACCCATCTTCTCC 7200
QY 2401 CysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu 2420
DB 7201 TGGCAATTCGGATACACACATGGGGCTCCAGGGTCCGTAAGTGCATGGCCATGGGCTC 7260
QY 2421 TyrSerGlySerGluValArgCysLeuAlaGlyHisCysGlyTyrProGluProIleVal 2440
DB 7261 TGGAGTGGCTGTGAAGTCCGCTGCTTGTGAGCACTGTGGAGACTGTGAGCCCATTTGTC 7320
QY 2441 AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsn 2460
DB 7321 AACGACACATCAATGGGAGAACTACAGCTACCCGGGACAGTGGTGTACCAATCAAT 7380
QY 2461 AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTyrPheSer 2480
DB 7381 GCTGGCTTCGAGCTGTATGGGACATGTCTGTGGCATCTGTCCACAGGATCATCATCTGGTG 7440
QY 2481 GlyLysThrProPheCysValProIleThrCysGlyHisPProGlyLysProValAsnGly 2500
DB 7441 GGCAGAACCCCTTCTGTGTGGCAATTACTGTGACACCCAGCAACCTGTCAAGGCG 7500
QY 2501 LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValCysAsnProGly 2520
DB 7501 CTCACACAGGTTACCAAGTTTAACTTCAACGATGTGTGCATAGTTTGTGTGCAACCTGGG 7560
QY 2521 TyrMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSerGlyGlnTyrPheSerPheMet 2540
DB 7561 TATATGGCTGAGGGGGGCTGTAGTCCATGCTGGGCCAGGSCAATGAGTACATG 7620
QY 2541 LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluAsnSerValArg 2560
DB 7621 CTGCCACCTGCAGATCATCACTGTACAGATCTGTGACACCAAGAAATATGTCTGCT 7680
QY 2561 GlnValHisAlaSerGlyProHisArgPheSerPheGlyTyrThrValSerTyrArgCys 2580
DB 7681 CAGGTTCACAGCCAGGCGCCGACAGTTCAGTTCGACACTGCTGTACCGGCTGCG 7740
QY 2581 AsnHisGlyPheTyrLeuLeuGlyTyrProValLeuSerCysGlnGlyAsnGlyThrThr 2600
DB 7741 AACCAAGGCTTACTCTCTGGGACCCCAAGTGTCAAGTGTCAAGTGTCTTACCGGCTGCG 7760
QY 2601 AspaArgProArgProGlnCysLeuLeuValSerCysGlyHisPProGlyLysProHis 2620
DB 7801 GACGTCTCCCGCCCAAGTGTCTGTGGTCTGTGGCCATCCGGGCTCCCGGCTCAC 7860
QY 2621 SerGlnMetSerGlyLysSerTyrThrValGlyAlaValAlaArgTyrSerCysIleGly 2640
DB 7861 TCCAGATGTGTGAGACAGTATATCTGTGGAGACAGTGTGGCTTACAGTGTGATCGGC 7920
QY 2641 LysArgThrLeuValGlyLysSerThrArgMetCysGlyLeuAspArgLysIleTyrThrGly 2660
DB 7921 AAGGCTACTGTGGGAGAAACAGCACCGCATGTGTGGGCTGATGACACTGACCTGGC 7980
QY 2661 SerLeuProHisCysSerGlyThrSerValGlyValCysGlyLysProGlyIleProAla 2680
DB 7981 TCCCTCCCTCACTGTCAAGAAACAGCGTGGAGTGTGGGATGCTCCGGGATCCGGCT 8040
QY 2681 HisGlyIleArgLeuGlyLysSerPheAspProGlyTyrThrValMetArgPheSerCysGlu 2700
DB 8041 CATGGCATCCGTTTGGGGGACAGCTTGTATCCAGGCACTGTGATGCCCTTCAAGCTGTGAA 8100
QY 2701 AlaGlyHisValLeuAspArgLysSerSerGluArgThrCysGlnAlaAsnGlySerTyrPheSer 2720
DB 8101 GCTGGCAACGCTCCCGGGATGCTCAGAGCGCACCTGTCAAGCAATGAGCTCGTGGAGC 8160
QY 2721 GlySerGlnProGluCysGlyValIleSerCysGlyLysAsnProGlyLysProSerAsnAla 2740
DB 8161 GGCTCGAGGCTGATGTGAGTGTATCTTGTGGGAACCTTGAGCTCCAAATATATGCC 8200
QY 2741 ArgValAlaPheSerAspGlyLeuValPheSerSerSerIleValTyrGluCysArgGlu 2760
DB 8221 CGAGTGTGTTCAGTGTATGGCTGTGTTTCTCAAGCTTATCTGTATGATGTCGGGAA 8280
QY 2761 GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyThrTyrThrGly 2780
DB 8281 GGAATCTACGCCACAGGCTGCTCAGCCGCTCATCTGCTCAATGATGATCTGTGACAGGC 8340
QY 2781 SerAspProGluCysLeuValIleAsnCysGlyLysAspProGlyIleProAlaAsnGlyLeu 2800
DB 8341 AGTGAACCTGTAGTCTCTGTGATTAATCTGTGTGACCTTGGGAAATTCACGCAATGGCTTT 8400
QY 2801 ArgLeuGlyAsnAspPheArgTyrAsnLysThrValThrTyrGlnCysValProGlyTyr 2820
```


QY 401 AlaProValLeuGlyThrPheSerGlyValAsnGlnLeuProSerSerIleThrSerSerGly 420
 Db 1201 GGGCCGCTCTGGGACCTTCTCAGAAACACAGCTTCCCTCCATCACAAGAGGTGC 1260
 QY 421 HisValAlaArgLeuGlnPheGlnThrAspHisSerThrGlyValArgGlyPheAsnIle 440
 Db 1261 CACGTGGCCCGCTCGAGTTCAGACTGACCTCCACAGGGAAGAGGGGCTTCAACATC 1320
 QY 441 ThrPheThrThrPheArgHisAsnGlnCysProAspProGlyValProValAsnGlyLys 460
 Db 1321 ACTTTTACACCTTCGACACACACAGATGCCCGGATCCTGGCGTTCAGTAATGGCAAA 1380
 QY 461 ArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGlnGlyPhe 480
 Db 1381 CGGTTGGGGACAGCTCCAGCTGGGACGCTCCATCTCTCTGTGATGAAGGCTTC 1440
 QY 481 LeuGlyThrGlnGlySerGlnThrIleThrCysValLeuLysGlnGlySerValValTyr 500
 Db 1441 CTTGGGACTCAGGGCTCAGAGACCATCACTCCGCTCGAAGAGAGGACGCTGCTGG 1500
 QY 501 AsnSerAlaValLeuArgCysGlnAlaProCysGlyGlyHisLeuThrSerProSerGly 520
 Db 1501 AACAGCGCTGTGCTGGGTGTGAAGCTCCGTGGTGTCACTGACTTCCGCCAGCGGC 1560
 QY 521 ThrIleLeuSerProGlyTyrProGlyPheTyrLysAspAlaLeuSerCysAlaTyrVal 540
 Db 1561 ACCATCTCTCTCCGGGCTGGGCTGGCTTCTTACAAAGATGCCGTGAGCTGGCGGTG 1620
 QY 541 IleGlnAlaGlnProGlyTyrProIleLysIleThrPheAspArgPheLeuThrGlnVal 560
 Db 1621 ATTGAGGCCACGACGACGCTACCCCATCAAAATCACTTCGACAGATTCAAAACGAGGT 1680
 QY 561 AsnTyrAspThrLeuGlnValArgAspGlyArgThrTyrSerAlaProLeuIleGlyVal 580
 Db 1681 AACATATGACACCTCGGAAGTACGCGATGGGCGGACTTACTCAGGCCCTTATCGGGGT 1740
 QY 581 TyrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAsnTyrLeuThrLeuLeu 600
 Db 1741 TACACGGGACCCAGAGTCCCGAGTCTCTCATCAGACACGACCAATCACTTACTCTCTC 1800
 QY 601 PheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgTyrGlnThrIleThr 620
 Db 1801 TTCTCTACGACAGAGAGTCACTCGGACATCGGCTTCCAGCTCCGTAAGAATCAATACA 1860
 QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn 640
 Db 1861 CTGCAGTCAAGACCACTGTCTGGATCCAGGAATCCAGTAATGACAGCGTCATGGGAT 1920
 QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660
 Db 1921 GACTTCTACGGGGGCGCTGAGTCACTTCAAGCTGTGACTGGGCTTACACTTAAGTAC 1980
 QY 661 GlyLysProLeuGlnCysGlnProAspPheGlnThrPheSerArgAlaLeuProSerCysGln 680
 Db 1981 GGGAGAGCTCTGAGAGTGTGAAGCCCACTTCAAGTGAAGCCGGGCTTCCAGATTGTGA 2040
 QY 681 AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
 Db 2041 GCTCTGTGTGGCTTCAATCAAGGCTCCAGTGGACATCTTGTGCCAGGGTTCCT 2100
 QY 701 AspPheTyrProAlaAsnLeuAsnCysThrTyrIleIleGlnThrSerHisGlyLysGly 720
 Db 2101 GACTTCTACCCCAACTTAACTGACCTGAGATTATCGAAATCAATCAGGAGAGGT 2160
 QY 721 ValPhePheThrPheHisThrPheHisLeuGlnSerGlyHisAspTyrLeuLeuIleThr 740
 Db 2161 GTGTTCTTCACTTCCACACTTCCACTGAAAGTGGCCATGACTACTCTCATCACT 2220
 QY 741 GlnAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
 Db 2221 GAGAAAGGACGACTTCAACCCAGCCCTGAGGAGAGTAAGTGAATCTGGCTGCCAGCTCC 2280

QY 761 IleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
 Db 2281 ATCAGCGCTGGAGCTCTATGGAACTTCACTCCCAAGTCCGCTTCATCTTGATTTCTCC 2340
 QY 781 MetSerTyrGlnGlyPheAsnIleThrPheSerGlnTyrAspLeuGlnProCysGlnGln 800
 Db 2341 ATGTCAATGAAGATTCAACATCACTTCCAGAGTACGACTTGGAGCCCTGTGAGAGAG 2400
 QY 801 ProGlnValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeu 820
 Db 2401 CCCAGAGTCCAGGCTTACAGCATCCGGAAGGCTTGGACGTTTGGCGTGGGACACCTTG 2460
 QY 821 ThrPheSerCysPheProGlyTyrArgLeuGlnGlyThrAlaArgIleThrCysLeuGly 840
 Db 2461 ACCTTCTCTCTTCCCGGGTACCGTGTGAGGGCACCGCCCATCACTGCTGGGG 2520
 QY 841 GlyArgArgArgLeuThrPheSerSerProLeuProArgCysValAlaGlnCysGlyAsnSer 860
 Db 2521 GGCAAGCGGGCGCTGTGGAGCTCCGCTTCCCAAGTGTGTGTGAGTGGGAATTCA 2580
 QY 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn 880
 Db 2581 GTCAAGGCACTCAGGGTACTTGTCTGTCCCAACTTTCCTGTGAACATAAACAAT 2640
 QY 881 HisGlnCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysAlaArg 900
 Db 2641 CATGAATGATCATCTCATCAGACACGACGAGGAAAGGAAATTCAGCTGAAGACGAGG 2700
 QY 901 AlaPheGlnLeuSerGlnGlyAspValLeuLysValTyrAspGlyAsnAsnAsnSerAla 920
 Db 2701 GCATTTCGAACCTTCGAAGAGAGATGCTCTCAAGTTTATGATGGCAACACAACTCCGCC 2760
 QY 921 ArgLeuLeuGlyValPheSerHisSerGlnMetMetGlyValThrLeuAsnSerThrSer 940
 Db 2761 CGTTTGTGGAGTGTTTTATGACCATTCGAGATGTAGGGGTGACTTTGAACAGACATCC 2820
 QY 941 SerSerLeuThrLeuAspPheIleThrAspAlaGlnAsnThrSerLysGlyPheGlnLeu 960
 Db 2821 AGCAGTCTGTGGCTTGAATTCATCACTGATGCTGAACACACAGCAAGGCTTTGAACGTG 2880
 QY 961 HisPheSerSerPheGlnLeuLeuLysCysGlnAspProGlyThrProLysPheGlyTyr 980
 Db 2881 CACTTTCCAGCTTGAATCATCAATATGAGAGACCCAGAAACCCCAAGTTTGGCTAC 2940
 QY 981 LysValHisAspGlnGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
 Db 2941 AAGTTCATGATGAAGGTCAATTTTGCAGGAGGCTCCGTCCTTCACTGAGACCTTGA 3000
 QY 1001 TyrSerLeuArgLysSerGlnGlnLeuLeuCysLeuSerGlyGlnArgTyrThrAsp 1020
 Db 3001 TACAGCTGTGGGGTATGAGAGAGCTGTGTGTGAGTGAAGAGCCCGGACCTTGGGAC 3060
 QY 1021 ArgProLeuProThrCysValAlaGlnCysGlyGlyThrValArgGlyGlnValSerGly 1040
 Db 3061 CGGCTCTGCCCACTGTGTCCCGAGTGTGAGAGGACAGTGAAGAGAGAGTGTCCGGGG 3120
 QY 1041 GlnValLeuSerProGlyTyrProAlaProTyrGlnHisAsnLeuAsnCysIleTyrThr 1060
 Db 3121 CAGGTGTGTCAACCGGGGTATCCAGCTCCATATGAACAAATTCACATCATGTGAGAC 3180
 QY 1061 IleGlnAlaGlnAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGlnGln 1080
 Db 3181 ATCGAAGCAGAGCGGGCTCAGACCATTTGGGTACACTTCTGGTGTGACACAGAGAGAG 3240
 QY 1081 ValHisAspValLeuArgIleTyrAspGlyProValGlnSerGlyValLeuLeuLysGln 1100
 Db 3241 GTTTCAGACAGTGTGGCCACTTGGAGTGGGCTGTGAGAGCCGGGTTCTGTGAGAGAG 3300
 QY 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValLeuGln 1120
 Db 3301 CTGAGTGGCCCGGCGCTGACCAAGAGACCTCATAGCACTTCAACTGGGTGCTGTCAG 3360
 QY 1121 PheSerThrAspPheThrSerLysGlnGlyPheAlaIleGlnPheSerValSerThr 1140

```
Db 3361 TTGAGCAGTACTTTCACACGAGCAGGCGCTTGCCATTCAATTTTCAGTCCACA 3420
Qy 1141 AATThSerCyAsnAspProGly11eProGlnAsnGlySerArgSerGlyAspSerTrp 1160
Db 3421 GCACATCTCTGCAATACCTCGAGTCCCGCAGAAATGGAGTCCGAGTGGACAGTTGG 3480
Qy 1161 GUAAGAGlyAspSerThrValPheGlnCyAspProGly11eProGlnAsnGlySerArg 1180
Db 3481 GAAGCGGCGACTCCACAGTGTTCAGTGTGACCTGGCTACGCGCTGACAGGAGTGA 3540
Qy 1181 GUA11eSerCyAsnVal11eGlnAsnArgPhePheTrpGlnProSerProProThCys 1200
Db 3541 GAGATCAGCTGTGTGAAGATCGAAGAACAGTTCTTCTGGACGCCAGCCCAACATGC 3600
Qy 1201 11eAlaProCyGly11eGlyAsnLeuThrGlyProSerGlyVal11eLeuSerProAsnTrp 1220
Db 3601 ATCGCTCTCTGCGGGGAGACTCGACAGACCATCTGGAGTCATCTCTCACCAAAATTAC 3660
Qy 1221 ProGluProTrpProProGly11eGlnCyAsnTrpLysVal11eVal11eSerProAsnTrp 1240
Db 3661 CCAGAACCTTACCCGCGAGGAGAGTGTGACTGGAAAGTACCTCTCACACAGACTAC 3720
Qy 1241 Val11eAlaLeuVal11ePheAsn11ePheAsnLeuGluProGly11eAspPheLeuH11e 1260
Db 3721 GTCATCGCTCTGTGATTTTAACTTTTAACTTTTAACTGGAGGCTGTGACTTCTCTCATATC 3780
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeu11eGlySerPheTrpGlySerGlnLeuPro 1280
Db 3781 TACAGCGAGCGGAGCTCTCTACGCCCTTCTCATAGAAAGCTTCTTAAAGCTCTCCAGCTCCA 3840
Qy 1281 G11eArg11eGluSerSerAsnSerLeuPheLeu11ePheArgSerAspAlaSerVal 1300
Db 3841 GCGCGCATTTGAAGACAGACGACACAGCTCTTCTCGCTTCCGACGATGATCTGTG 3900
Qy 1301 SerAsn11eGlyPheVal11eAspTrpThrGlnAsnProArgGluSerCysPheAspPro 1320
Db 3901 AGCATGCTGCTCTGTGATTTTAACTTTTAACTTTTAACTGGAGGCTGTGACTTCTCTCATATC 3960
Qy 1321 GlySer11eLysAsnGlyThrArgVal11eGlySerAspLeuLysGlySerSerVal11e 1340
Db 3961 GGTTCATCAAGACGAGCAGACGCGGTGGGTCCGACCTGAGCTGGCTCTCTCGTACAC 4020
Qy 1341 TyrTrpCysH11eGly11eGlyVal11eGluGly11eThrSerThrLeuSerCys11eLeuGly 1360
Db 4021 TACTACTGCCACGCGGGCTACGAAATTGAGGGGCACTCGACCTGTGAGCTGCACTCGGGG 4080
Qy 1361 ProAspGlyLysProVal11eTrpAsnAspProArgProValCysThrAlaProCysGly11e 1380
Db 4081 CCGATGGGAAGCCCGTGTGGAACATCCCGGCGAGTGTGACACGCCCTGTGTGGGGA 4140
Qy 1381 GlnTrpVal11eGlySerAspGlyVal11eLeuSerProAsnTrpProGlnAsnTrpThrSer 1400
Db 4141 CAGATGTGGGTTCGAGCGAGTGTGTGTCCGCCCAATCACCCCAAGAACATCACACGT 4200
Qy 1401 G11eGln11eCysLeuTrpPheVal11eThrVal11eProLysAspTrpVal11eGluGlnPhe 1420
Db 4201 GGAACATGCTGCTGTGATTTTGTGACTGTGCCCCAAGGACTGTGTGTGTGGCCAGTTC 4260
Qy 1421 AlaPhePheH11eThrAlaLeuAsnAspVal11eGluVal11eAspGlyH11eSerGlnH11e 1440
Db 4261 GCCTTCTTTCACAGCGCCCTCAACGAGTGTGTGAGTTCACGACGCGCCACAGCCAGC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerH11eThrGlyGluSerLeuProLeuAlaThr 1460
Db 4321 TCGCGGCTCTCGCTGCTCTCTCGGCTCCATACAGGAGATTCATGCTGCTTGGCCACC 4380
Qy 1461 SerAsnGlnVal11eLeu11eLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheH11e 1480
Db 4381 TCCAAATCAAGTTCTCTTAAGTTACGCGCAAGGCTCTGACACAGCGGAGGCTTTCAC 4440
Qy 1481 PheVal11eGlnAlaVal11eProArgThrSerAlaThrGlnCysSerSerVal11eProGluPro 1500
Db 4441 TTGTCTACCAAGCGGTTCTCGAACACAGGCCACGAGTGCAGTCTGTGCGGAAACC 4500
Qy 1501 ArgTrpGlyLysArgLeuGlySerAspPheSerVal11eVal11eArgPheGluCys 1520
Db 4501 CGCTATGGCAAGGCTGTGGAGTGTCTCTGTGGTGGGGCCATGCTCGCTTCGATATCC 4560
Qy 1521 AsnSerGly11eVal11eLeuGlnGlySerProGlu11eGluCysLeuProVal11eProGlyAla 1540
Db 4561 AACTCGGCTATGCTCTGACAGGGGTGCGCAGAGATGAGTGTCTCTGTGCTGGGGCC 4620
Qy 1541 LeuAlaGlnTrpAsnVal11eSerAlaProThrCysVal11eProCysGly11eVal11eLeuThr 1560
Db 4621 TTGGCCAAATGGAATGTCACAGCCCAAGTGTGTGTGCTGTGTGAGGCAACTCACA 4680
Qy 1561 GUAArgArgGly11eThr11eLeuSerProGlyPheProGluProTrpLysAsnSerLeuAsn 1580
Db 4681 GAGCGGAGGGGACATCTCTGCTCCCGCTTCCAGAGCCGTACTTCAACAGCTTCAC 4740
Qy 1581 CysVal11eTrpLys11eVal11eProGluGlyAlaGly11eGln11eGlnVal11eSerPhe 1600
Db 4741 TGTGTGTGAAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800
Qy 1601 Val11eThrGlnGlnAsnTrpAspSerLeuGlnVal11ePheAspGlyAlaAspAsnThrVal11e 1620
Db 4801 GTGACAGACGACATCTGGAATCTGCTGGAGATTTGATGTGAGATTAACATGTAACC 4860
Qy 1621 MetLeuGlySerPheSerGly11eThrVal11eProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGAGTTTCTCAGAAACACGCTGCTGCTGTGAAACAGACCTCCACACAG 4920
Qy 1641 LeuTrpLeuH11ePheTrpSerAsp11eSerVal11eSerAla11eGlyPheH11eLeuGluTrp 1660
Db 4921 CTCTACTCTCACTTCTTACTCAGATATCAGCTATCTGACGTGCTTCCACTTGGAGTAC 4980
Qy 1661 11eThrVal11eGlyLeuSerSerCysProGluProAlaVal11eProSerAsnGlyVal11eThr 1680
Db 4981 AAAAGGTGGGCTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5040
Qy 1681 GlyGluArgTrpLeuVal11eAspVal11eSerPheGlnCysGluProGly11eVal11eLeu 1700
Db 5041 GCGAGCGCTACTGTGGAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100
Qy 1701 GlnGlyH11eAlaH11eSerCysMetProGly11eThrVal11eArgAspTrpAsnTrpProPro 1720
Db 5101 CAGGGCAGCGCCACATCTCTGCTGATCCCGGAAACATGTGGGATGAATCACTTCTCT 5160
Qy 1721 ProLeuCys11eAlaGlnCysGly11eThrVal11eGluGluGluGlyVal11eLeuSer 1740
Db 5161 CCACCTGTGATGTGACAGTGTGGGGAACAGTGGAGAGATGGAGGGGCTGTGTGTGTGTGT 5220
Qy 1741 ProGlyPheProGlyAsnTrpProSerAsnMetAspCysSerTrpLys11eAlaLeuPro 1760
Db 5221 CCCGCTTCCAGGCAACCTACCCAGTACATGATGATGCTCTCGGAAATATGACACTGCCCC 5280
Qy 1761 ValGlyPheGlyAlaH11eGln11ePheLeuAsnPheSerThrGluProAsnH11eAspTrp 1780
Db 5281 GTGGGCTTTGAGAGTCAATCAATCTCTGAACTTCTCACCCAGCCCAACACAGACTAC 5340
Qy 1781 11eGlu11eLeuArgAsnGlyProTrpGluTrpSerArgMetMetGlyVal11ePheSerGlySer 1800
Db 5341 ATGAAATATCGGAATGCCCCCTTATGAGACCGAGCCGATGAGAGATTCAGTGTGAAC 5400
Qy 1801 GluLeuProSerSerLeuLeuSerThrSerH11eGln11eThrVal11eTrpPheH11eSerAsp 1820
Db 5401 GAGCTTCCAAAGTCTCTCTCTTCAAGTCTCCAGAGAACACCGTGTATTTCCACAGCGAC 5460
Qy 1821 H11eSerGlnAsnArgProGlyPheLysLeuGluTrpGlnAlaTrpGluLeuGlnGluCys 1840
Db 5461 CACTCCAGATGTGGCAGAGATTCAGATGAGATTCAGGCTATGAATTCACAGAGTGC 5520
Qy 1841 ProAspProGluProPheAlaAsnGly11eVal11eArgGlyVal11eGly11eTrpAsnVal11eGln 1860
Db 5521 CCAGACCCAGAGCCCTTGTCCAAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5580
```

Qy	1861	SerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys	1880	Qy	2221	HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuThrPheSerGluAlaIle	2240
Db	5581	TCAGTGCCTTCGAGTGCCTCCGGGGATCAATTGACTGGCCACCTGCTTCACGGT	5640	Db	6661	CACAGCATGGCCATCTGTATCCCGGACCCCAAGGCTACCACTGTGGACGGAACCATC	6720
Qy	1881	GlnHisGlyThrAsnArgAsnTrpAspHisProLeuProLysCysGluValProCysGly	1900	Qy	2241	ProLeuCysGlnAlaLeuSerCysGlyLeuProGlnAlaProLysAsnGlyMetValPhe	2260
Db	5641	CAACATGGCACCAACCGGAATTGGGACCAACCCCTGGCCAAAGTGTGAAGTCCCTTGAGGC	5700	Db	6721	CTCTGTGTCAAGCTCTTCTGTGGGCTTCGTGAAGGCCCAAGAAATGAAATGTGTGT	6780
Qy	1901	GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer	1920	Qy	2261	GlyLysGlyThrValGlyThrLysAlaValTyrSerCysSerGlyGlyTyrHisLeu	2280
Db	5701	GGGAACATCACTTCTCCACGGGCACTGTACTCCCGGGGTTCCCTAGCCCTACTCC	5760	Db	6781	GGCAGGAGTACACAGGGGAAACCAAGCCGTGTACAGCTGAGTGAAAGGCTACCACTC	6840
Qy	1921	SerSerGlnAspCysValIleTrpLeuIleThrValProIleGlyHisGlyValArgLeuAsn	1940	Qy	2281	GlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuThrPheSerAsnArgAsn	2300
Db	5761	AGCTCCACGAGCTGTGTGGCTATCATCCGTGCCCATGGCCATGGCGTCCGCTCAAC	5820	Db	6841	CAGGCAAGCCCTAGGCCATCTGCAGAGTGTCTGGACACAGGCTATGGACCAACCGCAT	6900
Qy	1941	LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln	1960	Qy	2301	ValProProGlnCysValProValThrCysProAspLysSerSerIleSerValGlnHis	2320
Db	5821	CTCAGCTGCTGCAGACAGAGCCCTCTGGAGATTTCATCACCATCTGGGATGGGCCACAG	5880	Db	6901	GTCCACACAGTGTGTCCTGTGACTTGTCTGTATGTCACTGATGATGATGATGATGATGAT	6960
Qy	1961	GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLysThrValGlnSer	1980	Qy	2321	GlyArgTrpArgLeuIlePheGlnThrGlnThrGlnPheGlnAlaGlnLeuMetLeuIle	2340
Db	5881	CAACACACACACCGGCTCGGCGCTTCACCCGAGCATGGCCAGAAACAGTGCAGAGT	5940	Db	6961	GGCCGATGGAGGCTTATCTTTGAGACACAGTATCAGTTCAGGCCCCAGCTGATGCTATC	7020
Qy	1981	SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAlaThrGlyGlyIlePheAla	2000	Qy	2341	CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys	2360
Db	5941	TCATCCACACACGAGCTCGCTCAAGTTCACCGTATGACAGCAGAGGGGGATTTGGCC	6000	Db	7021	TGTGACCTTGCTATCTATCTGAGCCAAAGGATCATCCGCTGTACGGCCATGGCCAA	7080
Qy	2001	IleAlaPheSerAlaTyrProLeuThrLysCysProProProThrIleLeuProAsnAla	2020	Qy	2361	TrpSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle	2380
Db	6001	ATAGCTTCTCCGCTTATCCATCAACCAATGCCCTCTCCACACATCTCCCAAGGCC	6060	Db	7081	TGGAGCTCGGGGCTCTAGCCCACTGCGAATATCTCTGTGGAGAGCTCCGATT	7140
Qy	2021	GlyValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu	2040	Qy	2381	ProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSer	2400
Db	6061	GAACTCCTCACAGAGATGAAGATTCAATATAGTGCATGTAACCTACAGATGCCTC	6120	Db	7141	CCCCCAATGGCCACCGCATCGGAACACTCTGTCTACAGGGCCAAACAGCCATCTTCC	7200
Qy	2041	ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGln	2060	Qy	2401	CysAsnSerGlyTyrThrThrLeuValGlySerArgValArgLysCysMetAlaAsnGlyLeu	2420
Db	6121	CTGGCTTTCCTTATGAGGGAAATGAATTCTGACTGCAAACTTGGAACTTACTCTCAG	6180	Db	7201	TGCATTCGGAATACACTGTGTGGCTCCAGGGTGGTGTAGTGCATGGCCATGGGCTC	7260
Qy	2061	PheGluGlyProProProIleCysGlyValHisCysProThrAsnGluLeuLeuThrAsp	2080	Qy	2421	TrpSerGlySerGluValArgCysLeuAlaGlyHisCysGlyThrProGluProIleVal	2440
Db	6181	TTTGAAGGACCAACCCCGATATGTGAGTGCATCTCCAAACAAATGAGCTTCTACAGAC	6240	Db	7261	TGGAGTGGCTCTGAGTCCGCTGCTGTGAGACATGTGGACTCCTGAGCCATTGTC	7320
Qy	2081	SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys	2100	Qy	2441	AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsn	2460
Db	6241	TCCACAGGCGTGCATCCGAGCCAGAGCTAACCTGMAAGCTATCCCAAGTTCACAGCTGC	6300	Db	7321	AACGACACATCAATGGGAGAACTACAGCTACCGGGGAGTGTGTGTCATGCAATGCAT	7380
Qy	2101	SerThrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu	2120	Qy	2461	AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisIleTrpSer	2480
Db	6301	TCTTGGCTGTGAGAGTGGAGCCCGACTATTAACATCTCCCAACAGTGAATACTTCTC	6360	Db	7381	GCTGGCTTCGCTGATCGGATGTCTGTGCGCATCTGCGACAGAGATCACTGAGTGC	7440
Qy	2121	SerGluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu	2140	Qy	2481	GlyLysThrProPheCysValProIleThrCysGlyHisAspProGlyAsnProValAsnGly	2500
Db	6361	AGCAGAGACATATGATGATTTGAATTTTATGATGCTCATACAGACAGAGTCTCTG	6420	Db	7441	GGCAGACCCCTTCTGTGTGCCATTAACCTGTGGACACCAAGCAACCTGTCAAGGC	7500
Qy	2141	LeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer	2160	Qy	2501	LeuThrGlnGlyAsnGlnPheAsnLeuAspValIleLysPheValCysAsnProGly	2520
Db	6421	CTGAAGGCGCTCAGTGGGAATTACTAGCTCCCTGATGTTCACAGCTCAAGCAACTCT	6480	Db	7501	CTCAGTCAAGGTAAACAGTTTAACTCAAGATGTGTCAAGTTTGTTCACACCTCGGG	7560
Qy	2161	ValTyrLeuArgTrpSerSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyr	2180	Qy	2521	TyrMetAlaGluGlyAlaIleArgSerGlnCysLeuAlaSerGlyGlnTrpSerAspMet	2540
Db	6481	GTTGACCTGTGCTGTATCTGATCACGCCCTACATCGGAAGGGCTTCMAATCCGCTAT	6540	Db	7561	TATATGGCTAAGGGGCTGTAGGTCCCAATGCTGCGCACGGGCAATGAGTGAATG	7620
Qy	2181	SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr	2200	Qy	2541	LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnLeuAsnSerValArg	2560
Db	6541	TCAAGCCCTTACTGAGCTTCGCCAGGGCTCACTCCATGGCTTCACTTACGGCCAGAC	6600	Db	7621	CTGCCACCTCGAGAAATCACTGTACAGATCTGTGACACCAAGAAATATGAGTTCGT	7680
Qy	2201	SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly	2220	Qy	2561	GlnValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArgCys	2580
Db	6601	AGCACCCAGCCCGGGGCTCATCACTTGGCTGCAACGCGGCTACCGGCTGTGTGGGA	6660	Db	7681	CAGGTCCACGCCAGCGGCCGACAGGTTCAAGTTCAGGCTTCGCGACCACTGTGTCCGGTGC	7740
Qy	6601	AsnHisGlyPheThrLeuLeuGlyThrProValLeuSerCysGlnGlyAspGlyThrTrp	6660	Qy	2581		2600

```

Db      7741 AACCAAGGGCTTACTCTCTGGGACCCCAAGTCTAGTCCAGAGGAATGGCAATGG 7800
Qy      2601 AAspArgProArgProGlnCysLeuLeuValSerCysGlyValProGlySerProProHis 2620
Db      7801 GACCGTCCCGCCGCCCCAGTGTCTCTTGTTGCTCTGGCCATCCGGGGTCCCCGCTCAC 7860
Qy      2621 SerGlnMetSerGlyAAspSerThrThrValGlyValValArgTyrSerCysIleGly 2640
Db      7861 TCCCAATGTCTGGAGACAGTTATCTGTGGAGAGAGTGGTGTACAGCTGCACTCCGC 7920
Qy      2641 LysArgThrLeuValGlyAAspSerThrArgMetCysGlyLeuAAspGlyValSerThrGly 2660
Db      7921 AAGCGTACTGTGTGGGAAACAGACCCGCAATGTGGCTGGAGACATGGACTGGC 7980
Qy      2661 SerLeuProHisCysSerGlyThrSerValGlyValCysGlyAAspProGlyValProAla 2680
Db      7981 TCCCTCCCTCACTGCTCAGGAACAGCGTGGAGATTGCGGTGACCTGGGATCCCGCT 8040
Qy      2681 HisGlyIleArgLeuGlyAAspSerPheAAspProGlyThrValMetArgPheSerCysGly 2700
Db      8041 CATGGCATCCGTTTGGGGGACAGCTTTGATCCAGGACCTGTGATGCGCTTCAAGCTGTGA 8100
Qy      2701 AlaGlyHisValLeuArgGlySerSerGlnArgThrCysGlnAlaAAspGlySerTrpSer 2720
Db      8101 GGTGGCCACGTCTCCGGGGATCGTCAGAGCGCACTGTCAAGCCATGGCTGTGAGC 8160
Qy      2721 GlySerGlnProGlnCysGlyValIleSerCysGlyAAspProGlyThrProSerAAspAla 2740
Db      8161 GGCTCGCAGCTGAGTGTGAGTGTGATCTCTGTGGAAACCTGGGACTCCAAAGTAATGCC 8220
Qy      2741 ArgValValPheSerAAspGlyLeuValPheSerSerIleValTyrGlnCysArgGlu 2760
Db      8221 CGAGTTGTGTCAAGTATGCGCTGTTTTCACGCTCTATCGTCTATGATGTCGGGAA 8280
Qy      2761 GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAAspGlyThrTrpThrGly 2780
Db      8281 GGATCTACGCCACAGCGCTGTGTCAGCGCTCACTGCGTCAATGGTACTGAGACAGCG 8340
Qy      2781 SerAAspProGlnCysLeuValIleAAspCysGlyAAspProGlyIleProAlaAAspGlyLeu 2800
Db      8341 AGTGACCTGAGTCTCGTCAATAACTGTGTGACCTGGGATTCCAGCAATGGCCTT 8400
Qy      2801 ArgLeuGlnAAspPheArgTyrAAspLeuThrValThrTyrGlnCysValProGlyTyr 2820
Db      8401 CGGCTGGGCATATGCTTCAGGTACCAACAAATGTGTCAATAGTGTGCTGCTGCTAT 8460
Qy      2821 MetMetGlnSerHisArgValSerValLeuSerCysThrTyrAAspArgThrTrpAAspGly 2840
Db      8461 ATGATGAGTCACTAGAGTATCTGTCTGAGCTGCAACAGGACCGGACATGGAAATGA 8520
Qy      2841 ThrLysProValCysValAlaLeuMetCysValAAspProProLeuIleProAAspGlyLys 2860
Db      8521 ACCAAGCCCTGTCGCAAGCTCTCATGTGCAAGCACTCCGCTATCCCAATGGGAAG 8580
Qy      2861 ValValGlySerAAspPheMetTrpGlySerSerValThrTyrAlaCysLeuGlnGlyTyr 2880
Db      8581 GTGTGGGCTCTCATCTTCATGTGGGCTCAAGTGAATTAATGCTGCTGAGGAGGTAC 8640
Qy      2881 GlnLeuSerLeuProAlaValPheThrCysGlnGlyAAspGlySerTrpThrGlyValLeu 2900
Db      8641 CAGCTCTCCCTGCGCGGTGTTCACCTGTGAGGAAATGGGTCTTGAACCGGAAGAGCTG 8700
Qy      2901 ProGlnCysPheProValPheCysGlyAAspProGlyValProSerArgGlyAArgArgGlu 2920
Db      8701 CCTCAATGTTTCCCTGTGTTCTGCGGGAGTCTGTGATCCCTCCGCGGAGAGAGAG 8760
Qy      2921 AspArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuValLeu 2940
Db      8761 GACCGAGGCTTCTCTCAACAGTCAATGTGTCTCTCTCTCTGCAATCCCTCTGTGTGTG 8820
Qy      2941 ValGlySerProArgArgPheCysGlnSerAAspGlyThrTrpSerGlyThrGlnProSer 2960

```

```

Db      8821 GTGGGCTCTCCACGACGAGTTTTCAGTCAATGGGACATGGAGTGGACCCAGCCAGC 8880
Qy      2961 CysIleAAspProThrLeuThrThrCysAlaAAspProGlyValProGlnPheGlyIleGln 2980
Db      8881 TGCATGATCTCCAGCTTCAGACAGTGTGGAGCCCTGTGTGTCACAGTTTGGGATACAG 8940
Qy      2981 AAspAAspSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlnLysGlyTyr 3000
Db      8941 AACAAATTCAGAGGCTACACAGGTGGAGACACAGTCTCTTCCGTTGTAAAAAGAGTAC 9000
Qy      3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAAspLeuThrTrpSerGlyThrPro 3020
Db      9001 CTGCTTCAGGGCTCCACCAAGAGAGCTGCTCCCAAACTGACCTGAGTGGAAACCCCA 9060
Qy      3021 ProAAspCysValProHisHisCysArgGlnProGlnThrProThrHisAlaAAspValGly 3040
Db      9061 CTGATGTGTCTCCCAACCACTGACAGGACCCAGAGAGCCAAAGCATGCCAAGCTGGG 9120
Qy      3041 AlaLeuAAspLeuProSerMetGlyTyrThrLeuIleThrProAlaArgArgAlaSerPro 3060
Db      9121 GCCCTGATTTGCTCCCTCATGGGCTAACAGCTCATTACTCTGCGCAGAGAGGCTTCTCC 9180
Qy      3061 SerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGlyGlnAlaSerArg 3080
Db      9181 TCAGGGGTGGCTCCAGACACCGCACTGCAAGGCGAGTGGACGCTGGACAGGCAAGCCGC 9240
Qy      3081 ProSerAlaTrpArgSerGlyProValGlyAAspProSerThrLeuProGlySerHisArg 3100
Db      9241 CCATCTGCTGGAGGTCCGGGCCAGTGGGAGAACCATTAACACTGCCGGAGGACACCGC 9300
Qy      3101 SerProLysPro 3104
Db      9301 TCACCAAGCCT 9312

```

RESULT 5
 ADH71141
 ID ADH71141 standard; DNA; 9951 BP.
 XX
 AC ADH71141;
 XX
 DT 25-MAR-2004 (first entry)
 XX
 DE Human gene of the invention NOV4 SEQ ID NO:37.
 KW ds; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;
 KW anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
 KW obesity; diabetes; infectious disease; metabolic syndrome X;
 KW dyslipidaemia.
 OS Homo sapiens.
 PN WO2003102155-A2.
 PF 11-DEC-2003.
 XX
 PD 03-JUN-2003; 2003WC-US017430.
 XX
 PR 03-JUN-2002; 2002US-0385120P.
 PR 04-JUN-2002; 2002US-0385784P.
 PR 05-JUN-2002; 2002US-0386041P.
 PR 06-JUN-2002; 2002US-0386047P.
 PR 06-JUN-2002; 2002US-0386376P.
 PR 06-JUN-2002; 2002US-0386453P.
 PR 06-JUN-2002; 2002US-0386864P.
 PR 06-JUN-2002; 2002US-0387016P.
 PR 07-JUN-2002; 2002US-0386796P.
 PR 07-JUN-2002; 2002US-0386816P.
 PR 07-JUN-2002; 2002US-0386931P.
 PR 07-JUN-2002; 2002US-0386942P.
 PR 07-JUN-2002; 2002US-0386971P.
 PR 07-JUN-2002; 2002US-0387262P.
 PR 08-JUN-2002; 2002US-0296960P.

PR	10-JUN-2002	2002US-0387400P
PR	10-JUN-2002	2002US-0387550P
PR	11-JUN-2002	2002US-0387610P
PR	11-JUN-2002	2002US-0387625P
PR	11-JUN-2002	2002US-0387634P
PR	11-JUN-2002	2002US-0387668P
PR	11-JUN-2002	2002US-0387696P
PR	11-JUN-2002	2002US-0387702P
PR	11-JUN-2002	2002US-0387836P
PR	11-JUN-2002	2002US-0387859P
PR	12-JUN-2002	2002US-0387933P
PR	12-JUN-2002	2002US-03879934P
PR	12-JUN-2002	2002US-0387960P
PR	12-JUN-2002	2002US-0388022P
PR	12-JUN-2002	2002US-0388096P
PR	13-JUN-2002	2002US-0389123P
PR	14-JUN-2002	2002US-0389118P
PR	14-JUN-2002	2002US-0389120P
PR	14-JUN-2002	2002US-0389144P
PR	14-JUN-2002	2002US-0389146P
PR	17-JUN-2002	2002US-0389729P
PR	17-JUN-2002	2002US-0389742P
PR	18-JUN-2002	2002US-0389884P
PR	19-JUN-2002	2002US-0390006P
PR	19-JUN-2002	2002US-0390209P
PR	21-JUN-2002	2002US-0390763P
PR	17-JUL-2002	2002US-0396706P
PR	06-AUG-2002	2002US-0401628P
PR	09-AUG-2002	2002US-0402156P
PR	09-AUG-2002	2002US-0402256P
PR	09-AUG-2002	2002US-0402389P
PR	12-AUG-2002	2002US-0402786P
PR	12-AUG-2002	2002US-0402816P
PR	12-AUG-2002	2002US-0402821P
PR	12-AUG-2002	2002US-0402832P
PR	13-AUG-2002	2002US-0403448P
PR	13-AUG-2002	2002US-0403459P
PR	13-AUG-2002	2002US-0403531P
PR	13-AUG-2002	2002US-0403532P
PR	13-AUG-2002	2002US-0403563P
PR	13-AUG-2002	2002US-0406317P
PR	13-AUG-2002	2002US-0403617P
PR	26-AUG-2002	2002US-0406182P
PR	26-AUG-2002	2002US-0406355P
PR	27-AUG-2002	2002US-0406240P
PR	12-SEP-2002	2002US-0410084P
PR	20-SEP-2002	2002US-0412528P
PR	23-SEP-2002	2002US-0412731P
PR	30-SEP-2002	2002US-0414801P
PR	30-SEP-2002	2002US-0414839P
PR	30-SEP-2002	2002US-0414840P
PR	30-SEP-2002	2002US-0414954P
PR	09-OCT-2002	2002US-0417186P
PR	09-OCT-2002	2002US-0417406P
PR	23-OCT-2002	2002US-0420639P
PR	28-OCT-2002	2002US-0421156P
PR	31-OCT-2002	2002US-0422690P
PR	01-NOV-2002	2002US-0423130P
PR	05-NOV-2002	2002US-0423798P
PR	05-NOV-2002	2002US-0423798P
PR	12-NOV-2002	2002US-0425453P

PA (CURA-) CURAGEN CORP.
XX
PI Alabrook J^P, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
PI Catterton E, Chappoval AI, Cradtree-Bokor JR, Edinger SR, Ellerman K;
PI Etemberg S, Gangolli EA, Gerlich-VL, Gorman L, Gunther E, Guo X;
PI Guev VY, Herrmann JL, Ji W, Kékuda A, Li L, Liu X, MacDougall JR
PI Macleachlan T, Malynker UM, Mexick R, Millet I, Mishra VS;
PI Padgugan M, Patcurujan M, Pena CEA, Peyman JA, Raha D, Rastelli L;
PI Rieger DK, Rothenberg ME, Siocore P, Shenoy SG, Shimkets R;
PI Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M;
PI Zhong H;

XX	WP1: 2004-081935/08.
DR	P-PSDB; ADH71142.
XX	New NOVX polypeptides and nucleic acid molecules useful for preventing or
PT	treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
PT	obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX	
PS	Example 4; SEQ ID NO 37; 1880bp; English.
XX	
CC	The invention relates to a novel isolated polypeptide (NOVX). A
CC	polypeptide of the invention has cytostatic, immunomodulator,
CC	neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
CC	antihypaemic activity, and may have a use in gene therapy, and as a
CC	vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
CC	any of the 303 fully defined nucleotide sequences given in the
CC	specification. The polypeptide is useful in the manufacture of a
CC	medicament for treating a syndrome associated with a human disease. The
CC	polypeptide, polynucleotide and antibody are useful in diagnosing,
CC	treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,
CC	Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
CC	diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
CC	further used as hybridisation probes, in chromosome mapping, tissue
CC	typing, preventive medicine, and pharmacogenomics. The present sequence
CC	encodes a NOVX polypeptide of the invention.
XX	
SO	Sequence 9951 BP; 2118 A; 2382 C; 2697 G; 2154 T; 0 U; 0 Other;
XX	
Alignment Scores:	
Pred. No.:	0
Length:	9951
Percent Similarity:	16512.00
Matches:	3037
Conservative:	1
Best Local Similarity:	97.784
Mismatches:	2
Query Match:	97.754
Indels:	69
DB:	12
Gaps:	3
US-10-016-248-2 (1-3104) x ADH71141 (1-9951)	
QY	1 MetAlaGlyAlaIaPpPpPpAlaLeuLeuLeuPpPpCySerLeuLeuLeuSerAspCyCs 20
Db	1 ATGGCGGGCGCCCTCCCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
QY	21 AlAsSerAnGlnArghisSerValGlyValGlyPpSerGlyLeuValValGlyGlnLe 40
Db	61 GCTAGCAATAGAGGACACTCCGTGGGCGTAGACCTCCGAGCTAGTCAAGAACAAATT 120
QY	41 GluLeuLysSerArgGlyValLysLeuMetPpSerLysAspAsnSerGlnLysThSer 60
Db	121 GAGTTGAAAGCTCGAGGTGTGAAGCTGAAGCCCAAGCAACACGCAACAAACAGCTT 180
QY	61 ValLeuThGlnAlaGlyValSerGlnGlyHisAsnMetCyPpPpAspPpGlyTLePro 80
Db	181 GTGTTAACTCAGGTGTGTGTGTGTCCCAAGACATTAATGTGTCCAGACCTCGGATACC 240
QY	81 GluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
Db	241 GAAAGGGGCAAAAGCTAGGCTCGATTTCAGATTAGATCCAGGCTCCAGTTCACTGC 300
QY	101 AsnGluGlyTyrAspLeuGlnGlySerLysValGlyLeuTyrGlyMetLysValSerAspMet 120
Db	301 AACGAGGGCGATATCACTCGCAAGGGTCCCAAGCGGATCACTGTATGAAAGTAGAGGAATG 360
QY	121 PheAlaAlaIaTrpSerAspHisArgPpProValCysArgAlaArgMetCysAspAlaHisLeu 140
Db	361 TTGGCGCTCGAGCGACACACAGGCCAGTCTGCCAGGCCCGCATGTGTGATGCCACCTT 420
QY	141 ArgGlyPpSerSerGlyLeuLeuThSerPpAsnPheProLeuGlnTyrAspAsnAsnAla 160
Db	421 CGAGGCCCGCTCGGAGCATATCACTCCCAATTTCCCAATTCAGTATGACAAACATGCA 480
QY	161 HisCysValaTrpLeuLeuThraLalaLeuAsnProSerLysValLeuValaLpPheGlu 180
Db	481 CACGTGTGTGATATCTCAACAGCTCAACCTTCAGAGTATCAAGCTGCTCTTGAG 540

QY	181	GIUPEAEPLeuGIuArgLYTYrAspThrLeuThrValGIyAspGIyGLINAspGLY	200
Db	541	GAGTTCATTGGAGAGGGGCTATGACACCTGCAGCGTGGGTATGGTTCAGAGATGGG	600
QY	201	AspGInLysPTrValLeuLYrMetSerGIuAsnValaCySerAspSerProHsrPro	220
Db	601	GACCAAGAACAGATTCTTACATGCTTCMAAATCCCTGCAGTGCACACCTCCACACCCA	660
QY	221	GLYSerArgIIeProGIuSerMetSerGIyAspIIeTrpArgIIuysTrpThrValLeu	240
Db	661	GGCTCTCCACCCAGAGAGCATGCTGGGGACATCGAGGACAGAAATGACTTACTT	720
QY	241	GIuIIECySAAGAspIIeSerSerAspAlaArgSerGIySerValaArgIysSerPro	260
Db	721	GAGATCTGTCGTGACATTGACATTCAGATGCAGAGGACAGGTTCAGTGAAGAAAGTCTCCA	780
QY	261	LYrThrSerAsnAlaValGIuLeuValaIProGIyThrGIuIIEGIuGIuArgIysSerCys	280
Db	781	AAGACTTCATATGCTGTGGAACTGTGTGCTCTGGGACAGATTCAGACAGGGCACTTCC	840
QY	281	GLYAspProGIyIIeProAlaLYrGIyArgArgGLuGIySerArgPheHsrIsGIyAsp	300
Db	841	GGTACCTTGACATCTGCATATGGCCGAGGAGAAAGGCTCCCGGTTCACACCGGTGAC	900
QY	301	ThrLeuLysPheGIuCYsGInProAlaPheGIuLeuValGIuGIuLysAlaIIeThrCys	320
Db	901	ACACTCAAGTTTGATGATCCAGCCCGCTTTGACTGTGGAGACAGAAAGCAATCAATGC	960
QY	321	GIuLysAsnAsnGIuTrpSerAlaLYrLYrProGIyCYsValaPheSerCyAspPheAsn	340
Db	961	CAAAAGAAATACCAATGGCTGGCTTAAGAACCGAGCTGGTGTCTCGTCTTCAAC	1020
QY	341	PheThrSerProSerGIyValValIIEAspSerProAsnLYrProGIuAspLYrGIyAsnHsr	360
Db	1021	TTTACACAGCCCGTCTGGGTGTCTGCTCTTCCAACTACCCAGAGACTATGGCAACAC	1080
QY	361	LeuHICyValaTrpLeuIIeLeuAlaArgProGIuSerArgIIeHsrLeuAlaPheAsn	380
Db	1081	CTTCACGTGTCTGGCTCATCTCGGCGCAGGCTCGAAGCGCATCCACTGCGCTTCAAC	1140
QY	381	AspIIeAspValGIuProGIuPheAspPheLeuValaIIeLYrAspGIyAlaThrAlaGIu	400
Db	1141	GACATTGACGTGAGGCTCAGATTGATTTCTGTGTATCAAGAGAGGGGCAACCGCCAG	1200
QY	401	AlaProValLeuGIyThrPheSerGIyAsnGIuLeuProSerSerIIeThrSerSerGIy	420
Db	1201	GGCCCGGTCTGGGACCTTTCAGAGAAACAGCTTCCCTCTCATACACAAGCATGGC	1260
QY	421	HisValaIaArgLeuGIuPheGIuHsrAspHsrSerThrGIyLYrArgGIyPheAsnIIe	440
Db	1261	CACGTGGCCCGTCTGAGTTCAGACTACCACTTCACAGGAAAGAGGGGCTTCMAATC	1320
QY	441	ThrPheThrThrPheArgHisAsnGIuCYsProAspProGIyValaProValAsnGIyLys	460
Db	1321	ACTTTTACCACTTCCGACACACAGAGGCCCGGATCTGGCGCTTCCAGTAAATGGGAAA	1380
QY	461	ArgPheGIyAspSerLeuGIuLeuGIySerSerIIeSerPheLeuCYAspGIuGIyPhe	480
Db	1381	CGGTGTGGGAGAGCCTTCACGTGGGACGTCCATCTCTTCCCTGTGATGAAGGCTTC	1440
QY	481	LeuGIyThrGIuArgIysSerGIuThrIIeThrCYsValLeuLysGIuArgIysSerValaTrp	500
Db	1441	CTTGGGACTCAGGGGCTCAGAACCATCACTGCGCTTCGAAGAGAGGGACAGTGTCTGG	1500
QY	501	AsnSerAlaValLeuArgCYsGIuAlaProCYsGIyGIyHsrLeuThrSerProSerGIy	520
Db	1501	AACAGCGCTGTGCTGGGTGGAAGCTCCCTGTGTGTGTCACTTCGCCACAGGCGG	1560
QY	521	ThrIIeLysSerProGIyLYrTrpProGIyPheThrLYrAspAlaLeuSerCYsAlaTrpVal	540
Db	1561	ACATTCCTCTCTCCGGGCTGGCTTGACTTTCACAGAGATCCCTTGAAGCTGTGGCTGGGTG	1620

QY	541	ILEGLUALAGINPROGIYTYRPROLLELYNLETHPHEASAPRGPEUSTRGIUVAL	560
DB	1621	ATTTAGGGCCAGCCAGGCTACCCCATCAAAATACCTTCACAGATTCAAAACCGAGGTC	168
QY	561	ASNTYRASPTHRLEUGIULVALARGAASGLVARGTHRTRYSERALPROLEUILEGIYVAL	580
DB	1681	AACTATGACACCCCTGGAAAGTAAACCGATGGCGGACCTTACTACGCCCTTGATCGGGGATT	174
QY	581	TYTHISGLYTHRGINVALPROGINPHELEUILESETRHSEASNTYRLEUITYRLEULEU	600
DB	1741	TACACGGGAGCCAGGTTCCCGAGTTCCATACAGACAGCAACTACCTTACCTCC	1800
QY	601	PHSESTRHAPLYSSERHISSEASPILEGYPHGINLEUARGVTCULNTHRIETHR	620
DB	1801	TTCTCTACCGACAAGAGTCACTCGACATCGGCTTCACCTCCGCTATAGACTATAC	186
QY	621	LEUGINSEASPHISCYSELEUASPPROGIYILEPROVALASNGLYGINARHISGLYASN	640
DB	1861	CTGACAGACAGACCACTGCTGGATCCAGGAATCCCGATMAATGACAGGCTATGGAAAT	1922
QY	641	ASPPHETRYVALGIYVALALEUVALTHRPHSESTRYASPSERGLYTRYTHRLEUASERP	660
DB	1921	GACTTCTACGTGGGCGCGCTGGTAGACTTACACTGATCGGGCTAACATTAAGTAC	1980
QY	661	GLYGLUPROLEUGIUCYASGLUPROASNPHEGINTPSEARVALALEUPROSERCYSGIU	680
DB	1981	GGGAGACCTCTGGAGGTGAGCCCAACTTCAGTGAAGCGGGGCCCTCCAGTTGTAA	2040
QY	681	ALALEUCYSGIYGLYPHEIILEGINLYSERSEGLYTHRIILEUASERPPOGIYPHEPRO	700
DB	2041	GCTCTCTGAGTGAGCTTCAATTAAGCTCCAGTGGAACTCTGTCCCGAGGTTCCCT	2100
QY	701	ASPPHETRYPROASNAASNUASNCYSYTHTRPILEILEGULNTHRSEHISGLYVLSGLY	720
DB	2101	GACTTCTACCCCAACAACCTGAACTGCAGCTGGATTACGAAACATCTCATGGCAAGGCT	2160
QY	721	VALPHEPHETHRPHENISTHRPHENISLEUGIUSERGIYHISAAPTYRLEULEUILETHR	740
DB	2161	GGTGTCTTCACTTTCACACCTTCACCTGGAAAGTGCCCACTGACTCTCTCATCACT	2220
QY	741	GIUASNGLYSERPHETHRGINPROLEUARGIINEUPTHRGISYSTRAGLEUPROALAPRO	760
DB	2221	GAGAACGGAGCTTACCCACGACCCCTTAGAGCAGCACTACTGATCTCGGCTGCCAGCTCC	2280
QY	761	ILESERIALAGLYLEUITYRGIYASNPHETHRALAGINVALARGPHEIILESEASPPHESE	780
DB	2281	ATCACCGCTGGGCTCTATAGGCACACTTACCTCCAGAGTCGGCTTCATCTCGAATTTCTCC	2340
QY	781	MESESTRYRGIUGIYPHASNLIETHPHSESTRUITYRASPHEUGIUPROCYSGIUGIU	800
DB	2341	ATGTCAATGAAAGATTCAATCACTCTTTCAGAGTACATCTTGAGCCCTGTGAGAG	2400
QY	801	PROGIUVALPROALATYRSERIEARGVLSGLYLEUGINPHEGLYVALGIYASPTHLEU	820
DB	2401	CCCGAGGCTCCAGGCTACAGCATCCGGAAGGGCTTGCAGTTTGCGGCGGCGACACTTG	2460
QY	821	THRPHSESTRYSPHEPPOGIYTRYARGLEUGIUGIYTHRIALARGILETHRCYSEUGIY	840
DB	2461	AACCTTCTCTCTTCCCGGGGATCCGTGTGAAGGCGACCGCCGCACTACAGTCCCTGGGG	2520
QY	841	GLYARGARGARGLEUPTRPSESTRERPROLEUPROARGCYSEVALIAGIUCYSGIYASNSER	860
DB	2521	GGCAGACGGCGCTGTGGAGCTCGGCTCTGCAAGGTGTCTTCTAGTCTGGGAATTCA	2580
QY	861	VALNHRGLYTHRGINGLYTHRILEULEUSERPROASNPHEPROVALASNTYRANASNAEN	880
DB	2581	GTCACAGGACCTACGGGTACTTTCGTGTCCCGCACTTCTCTGTGAATCAATAAACAAT	2640
QY	881	HISGLIUCYSILETRYSERILEGINTHNTHGINPROGIYLVASGLYILEGLINLEUYSLAARG	900
DB	2641	CATGAATGATCTCACTCCATCCAGACCCAGCCAGGAAGGAATTCAGCTGAAAGCCAGG	2700
QY	901	ALAPHEGIULEUSERGIUGIYASPYVALLEUYSVALTYRASPGLYASNAASNAESERALA	920


```
Db 2701 GCATTCGAACCTCCGAGAGAGATGTCTCAAGGTTTATGATGGCAACAACATCCGCC 2760
Qy ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValIThrLeuAsnSerThrSer 940
Db 2761 CGTTTGCTGGAGATTTTAGCATTCGTAGATGATGGGGTGACTTTTGAAACAGACATCC 2820
Qy SerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerLysGlyPheGluLeu 960
Db 2821 AGCAGTCTGTGGCTTGAATTTATCATCTATCTGAAAACACAGACAGGGCTTTGAACCTG 2880
Qy HisPheSerSerPheGluLeuIleLysCysGluAspProGlyIThrProLysPheGlyThr 980
Db 2881 CACTTTTCCAGCTTTGAACCTCATCAATAGTAGAGACCAGAACCCCAAGATTGGCTAC 2940
Qy LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
Db 2941 AAGGTTCAATGATGAAGGTCATTTTGCAGAGGAGCTCCGTCTTCAGCTGTGACCTGGA 3000
Qy TyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerSerGlyLysArgThrTrpAsp 1020
Db 3001 TACAGCTCGCGGGGTAGTAGAGAGCTGTGTCTGAGTGAAGAGCCCGGACCTGGGAC 3060
Qy ArgProLeuProThrCysValAlaGluCysGlyIThrValArgGlyGluValSerGly 1040
Db 3061 CGGCTCTGCCACCTGTGTGCGCGAGTGTGAGAGGACATGAGAGAGAGTGTCCGGG 3120
Qy GluValLeuSerProGlyIThrProAlaProGlyIHisAsnLeuAsnCysIleTrpThr 1060
Db 3121 CAGGTGTCTCACCCGGGTATCCAGCTCCCTATGAAACAACTTCATCTGATCTGACCC 3180
Qy IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
Db 3181 ATCGAAGCAGAGGCCGCGCTGCACCATTTGGCTTACATCTTCTGTGTGTGACACAGAGAG 3240
Qy ValHisAspValLeuArgIleTrpAspGlyProValGluSerGlyValLeuLeuLysGlu 1100
Db 3241 GTTCACAGCGTGTGCGCATGTGGGATGGGCTGTGAGAGCGGGGTTCTGTAAAGAG 3300
Qy LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValIleLeuGln 1120
Db 3301 CTGATGGCCCGGCGCTGCGCCAGAGACCTGCATAGCACCTTCACCTCGTGTCTGAG 3360
Qy PheSerThrAspPhePheThrSerLysGlnGlyPheAlaIleGlnPheSerValSerThr 1140
Db 3361 TTCAGCACTGACTTCTTTCACAGACAGAGGAGCTTTGCCATTCAATTTTCAGTCCACA 3420
Qy AlaThrSerCysAsnAspProGlyIleProGluHisGlnLysSerArgSerGlyAspSerTrp 1160
Db 3421 GCAACGTCTCTGCANTGACCTCGGGATCCCGAGAAATGGGACTCGGAGTGTGACAGTTGG 3480
Qy GluAlaGlyAspSerThrValPheGlnCysAspProGlyIThrAlaLeuGlnGlySerAla 1180
Db 3481 GAACCGCGGCACTCCACAGTGTTCAGTGTGACCTCGCTACGGCTGCAGAGGAAGTGA 3540
Qy GluIleSerCysValLysIleGluAsnAspPhePheTrpGlnProSerProProThrCys 1200
Db 3541 GAGATTCAGCTGTGAGATGAGAAACAGGTTCTTCTGCGAGCCAGCCGCCCAACATGC 3600
Qy IleAlaProCysGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTrp 1220
Db 3601 ATGCTCCTCGCGGGGAGAGCTGACAGGACCATCTGGAATCATCTCTCCACCAATTAC 3660
Qy ProGluProGlyProProGlyLysGluCysAspTrpLysValIThrValSerProAspTrp 1240
Db 3661 CCAAAACCTTACCCGCGCAGAGAGAGTGTGACTGAAAGTGAACCGTCTCCACAGACTAC 3720
Qy ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyIThrAspPheLeuHisIle 1260
Db 3721 GTCAATGCGCTGTGATTTTAACTTTTAACTTGAAGCTGCGCTAAGACTTCTCCATATC 3780
Qy TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheIleGlySerGlnLeuPro 1280
Db 3781 TACGACGAGACGAGACTCTCTCAGCCCTCTCATAGGAAGCTTCTATGGCTCCAGCTCCA 3840
Qy GlyValGluIleGluSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GGCCGCAATTAAAGCAGACAGCAACAGCTCTCTCTCCCTTCCGACAGATGATCTGTG 3900
Qy SerAsnAlaGlyPheValIleAspTrpThrGluAsnProArgGluSerCysPheAspPro 1320
Db 3901 AGCAATCTGGCTTCGTCATCTAGTACTATACAGAAAACCCGGGAGATCATGTTTGAACCT 3960
Qy GlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSerValThr 1340
Db 3961 GGTTCCATCAAGAACGCGCACACGCGGTGGGTCCGACCTGAAGCTGGGCTCTCCGTACCC 4020
Qy TyrTrpCysHisGlyGlyIThrGluValGluGlyIThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTACTGCCACGCGGGGCTACGAAATGAGAGGACCTCGACCTCGACCTCATCTGAGG 4080
Qy ProAspGlyLysProValTrpAsnAspProArgProValCysThrAlaProCysGlyGly 1380
Db 4081 CTGATGGGAAGCCCGTGTGGAACAATCCCGGCGCACTGCACAGCCCTGTGGGGGA 4140
Qy GlnTrpValGlySerAspGlyValIleLeuSerProAsnTrpProGluAsnTrpThrSer 1400
Db 4141 CAGTATGGGTTTCGACGAGATGTCTTGTCCCCCACTAACCCCGAAGACTACACAGT 4200
Qy GlyGlnIleCysLeuTrpPheValIThrValProLysAspTrpValValPheGlyGlnPhe 1420
Db 4201 GGACAGATCTGCTGTGATTTTGTATCTGTGCCAAGGACATGATGAGTGTGGCAGATTCC 4260
Qy AlaPhePheIleThrAlaLeuAsnAspValAlaGluValHisAspGlyHisSerGlnHis 1440
Db 4261 GCCTCTTTTACACAGCGCCCTCAACAGAGCTGTGTGAGTTTACACAGGCCACAGCAGAC 4320
Qy SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeuAlaThr 1460
Db 4321 TCGGAGCTCTCAGCTCCCTCGGGGCTCCCATACAGAGAAATCATCGCTTGGCCACC 4380
Qy SerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGlyPheHis 1480
Db 4381 TCCAAATCAATTTCAATTAAGTTCAAGCGCCAAAGGCTTCACACAGCAAGGCTTCCAC 4440
Qy PheValTrpGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
Db 4441 TTTGTCTTACCAAGCGTTCCTCGAACAGGCGCCACAGAGTGCACCTGTGTGCCGAACCC 4500
Qy ArgTrpGlyLysArgLeuGlySerAspPheSerValGlyAlaIleValArgPheGluCys 1520
Db 4501 CGTATGGAAGAGGCTGGGCAAGTACTTCGGTGGGGGCCATTCGCTCGAATGC 4560
Qy AsnSerGlyTrpAlaLeuGlnGlySerProGluIleGlyCysLeuProValProGlyAla 1540
Db 4561 AACTCCGCTATGCTCCGCGAGGGGTCCGAGAAATCAGAGTCCCTCGTGTGCGTGGGCC 4620
Qy LeuAlaGlnTrpAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr 1560
Db 4621 TTGGCCCAATGAAATGTCTTACGGGCCCACTGTGTGTGTCGCTGTGAGGCAACTTACA 4680
Qy GluArgArgGlyIleThrIleLeuSerProGlyPheProGluProTrpTrpLeuAsnSerLeuAsn 1580
Db 4681 GAGGCGAGGGGCAACATCTGTCCCTCGCTTCCAGAGCGTAACTCAACAGCTCAAC 4740
Qy CysValTrpLysIleValValProGluGlyAlaGlyIleGlnIleGlnValIleSerPhe 1600
Db 4741 TGTGTGGAAGATCGGTGCTCCCGAAGGGCGTGGCATTCAGATCCAAAGTTGTCAGATTTC 4800
Qy ValThrGluGlnAsnTrpAspSerLeuGluValPheAspGlyAlaAspAsnTrpValThr 1620
Db 4801 GTGACAGAGCAAGAACTGGACTCGCTGGAAGTATTGATGTGTGACATTAACCTGTAAACC 4860
Qy MetLeuGlySerPheSerGlyIThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGAGATTCTCAGGAACAACGTCGCTGCTTGTGAACAGCACTCCAAACAG 4920
```

Qy	1641	LeuTyrLeuHisPheTyrSerAspIleSerValSerAlaAlaGlyPheHisIleuGluTyr	1660
Db	4921	CTGTACTTCATTTCTACTACATATACAGATATCTGACGTGGCTTCCACTTGGAGTAC	4980
Qy	1661	LysThrValAlaGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValIleYsthr	1680
Db	4981	AAATCGGTGGGCTGAGCAGATTGTCCGGAACCTGGTGTCCAGTACCGGGTGAAGCT	5040
Qy	1681	GLYGLARGTyrLeuValAsnAspValIleSerPheGlnCysGluProGlyTyrAlaLeu	1700
Db	5041	GGGAGCGCTACTTGTGATGATGTGTCTTTCAGTGTGACCCGGATATGCTCTC	5100
Qy	1701	GlnGlyHisAlaHisIleSerCysMetProGlyTyrThrValArgAspTyrAsnTyrPro	1720
Db	5101	CAGGGCCAGCCCACTCTCCGTCGATGCCGGAACAGTGGCGCATGGAATCACTCTCT	5160
Qy	1721	ProLeuCysIleAlaGlnCysGlyGlyThrValGluGluMetGluGlyValIleLeuSer	1740
Db	5161	CCACTCTGTATTGCAACAGTGTGGGGAAACAGTGAGAGATGAGAGGGGTGATCTGAGC	5220
Qy	1741	ProGlyPheProGlyAsnTyrProSerAsnMetAspCysSerTyrIleAlaLeuPro	1760
Db	5221	CCGGGCTTCCAGGAACTACCCCACTGATACATGAGCTGCTCTGGAAATATGACACTGCC	5280
Qy	1761	ValGlyPheGlyValHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspTyr	1780
Db	5281	GTGGGCTTGGAGCTCACATCCAGTCTCTGAATCTTCCACCGAGCCCAACACGATAC	5340
Qy	1781	IleGluIleArgAsnGlyProTyrGluThrSerArgMetMetGlyValPheSerGlySer	1800
Db	5341	ATAGAAATCCGGAAATGGCCCTTATGAGCCAGCCGATGATGGGAAATTCAGTGGAGC	5400
Qy	1801	GluLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTyrPheHisSerAsp	1820
Db	5401	GAGCTTCCAACTCCCTCTCTCCACGTCACAGACCAACCGTATTTTCCACAGCGAC	5460
Qy	1821	HisSerGlnAsnArgProGlyPheValLeuGluTyrGlnAlaTyrGluLeuGlnGlyCys	1840
Db	5461	CACCTCCAGATCGGACCGCAAGATTCAGCTGAGATCAGGCTATGAACTTCAAGAGTGC	5520
Qy	1841	ProAspProGluProPheAlaAsnGlyIleValArgGlyValArgIleValIleGln	1860
Db	5521	CCAGATCGCACCAACCGAATCTGGCAATGTGTGAGGGAGACTGCTACCAACGCGGACAA	5580
Qy	1861	SerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys	1880
Db	5581	TCACTGACTTTCAGTGTCTCCCGGGGTATCAATTGACTGGCCACCTGTCTCAGTGT	5640
Qy	1881	GlnHisGlyThrAsnArgAsnTyrAspHisProLeuProLysGlyValProCysGly	1900
Db	5641	CAACATGGCACCAACCGAATCTGGACACCCCTGCCCCAAGGTGAAATCTCTGTGTGC	5700
Qy	1901	GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer	1920
Db	5701	GGGAACATCACTTCTTCCAAAGGCACTGTGTACTCCCGGGGTTCCTTACGCCGTACTCC	5760
Qy	1921	SerSerGlnAspCysValTyrLeuIleThrValProIleGlyHisGlyValArgLeuAsn	1940
Db	5761	AGCTCCAGACACTGTGTCTGTGATACCGGTCCCATTTGGCCATGGCTCCGCTCAAC	5820
Qy	1941	LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyProGln	1960
Db	5821	CTCAGCTCTGTGACAGAGGCTCTGTGAGATTTTCAATCAATCTGGAGTGGGCCAAG	5880
Qy	1961	GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLeuThrValGlnSer	1980
Db	5881	CAAAACGACCAACGCTGTGGCGCTTTCACCCCGAGACATGGCCAAAGAAAACGTGACAGT	5940
Qy	1981	SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaIleThrGlyGlyIlePheAla	2000
Db	5941	TCATCCAAACCAAGTCTCTGCTCAAGTTCCACGTGATGACCAAGGGGGGATCTTGTGCC	6000
Qy	2001	IleAlaPheSerAlaTyrProLeuThrLysCysProProThrThrIleLeuProAsnAla	2020
Db	6001	ATAGCTTCTCCGCTTATTCACATCCACCAATGCCCTCTCCACCAATCTCCCAACGCGC	6060
Qy	2021	GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu	2040
Db	6061	GAAGTGTGTCAACAGATGATGAGAAATTCATATATAGTGCATGTGATGATGATGCTTC	6120
Qy	2041	ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysValLeuGlyThrTyrLeuGln	2060
Db	6121	CTGGCTTATACCTTATGTGGGAAATGAAATTTCTGACCTGCAAACTTGGAACTTACGACG	6180
Qy	2061	PheGluGlyProProProIleCysGluValHisCysProThrAsnGluLeuLeuThrAsp	2080
Db	6181	TTTGAAGACCAACCCCGATATGTGATGTGACCTGTCCAAACAAATGAGCTTGTGACAGC	6240
Qy	2081	SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys	2100
Db	6241	TCCACAGGCGGTATCTGTGACCAAGACTACCTGGAAAGCTATCCCACTTCCAGACTGC	6300
Qy	2101	SerTyrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu	2120
Db	6301	TCTTGGCTGTGAGAGTGGAGCCGACATTAACATCTCTCCACAGTGAATCTCTCTC	6360
Qy	2121	SerGluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu	2140
Db	6361	ACCGAAGAACATATGATGAGATTTGAGATTTTTCATGTGTCATGACAGACAGATCTCTG	6420
Qy	2141	LeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValIleThrSerSerAsnSer	2160
Db	6421	CTGAAGCCCTCAGTGGGAATTACTCACTCCCTCGATTTGACACAGCTCAAGCAACTCT	6480
Qy	2161	ValTyrLeuAlaGlyTyrSerSerAspHisAlaTyrAsnArgIleGlyPheLysIleArgTyr	2180
Db	6481	GTGATCCTGCGTGTGATCTGATCAAGCTTCAATCAATGGAAGGCTTCAAGATCGCTAT	6540
Qy	2181	SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr	2200
Db	6541	TCA-----	6543
Qy	2201	SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly	2220
Db	6543	-----	6543
Qy	2221	HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisIleuTyrSerGluAlaIle	2240
Db	6543	-----	6543
Qy	2241	ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPhe	2260
Db	6544	-----GCTCTTCTGTGGCTTCTGAGAGCCCCCAAGATGGAATGTGTCTT	6591
Qy	2261	GlyLysGluTyrThrValGlyThrLysAlaValTyrSerCysSerGluGlyTyrHisIleu	2280
Db	6592	GGCAAGAGATACACAGTGGGAAACCAAGGCCATGTACAGTGCAGTAAAGCTTACCACTC	6651
Qy	2281	GlnAlaGlyValaGlnAlaThrAlaGluCysLeuAspThrGlyLeuThrPheSerAsnArgAsn	2300
Db	6652	CAGGCAAGGCTTACAGGCCACTGACAGTGTGTGAAACAGGCTTATGAGCAACCGCAAT	6711
Qy	2301	ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGluHis	2320
Db	6712	GTCCCAACCAAGTGTCTCTGTGACTTGTCTGTGATGTCAATGATCATCGCTTGAGCAT	6771
Qy	2321	GlyArgTyrArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle	2340
Db	6772	GGCCGATGAGGCTTATCTTTAAGACACAGTATCAAGTTCCAGGCCCACTGATGTCTCATC	6831
Qy	2341	CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLys	2360
Db	6832	TGTGACCTGTGGCTACTATACTGTGCCAAAGGATCATCCGTGTCAAGGCCCAATGGCAAA	6891
Qy	2361	TyrSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle	2380

Db 6892 TGGAGCCCTGGGGAGCTACGCCCACTGGCCGAATCATCTCTGGAGAGCTCCGATT 6951
 Qy 2381 ProProAenGIyHLeaRgIleGIyThrLeuSerValIyRgIyAlaThrAlaIlePheSer 2400
 Db 6952 CCCCCCAATGGGCAACCGGATCGGAACACTGTCTGTACGGGGCAACCACTCTTCTCC 7011
 Qy 2401 CysAenSerGIyTyRThrLeuValGIySerArgValArgGluCysMetAlaAenGIyLeu 2420
 Db 7012 TGCATTCGGATACACACTGGTGGGCTCCAGGGTGGCTGTAGTGCATGGCCAAATGGGCTC 7071
 Qy 2421 TrpSerGIySerGIyValArgCysLeuAlaGIyHisCysGIyThrProGIyLeuAla 2440
 Db 7072 TGGATGGCTCTGAAGTCCGCTGCTGTGGACACTGTGGGAGCTCCGAGCCCATTTGTC 7131
 Qy 2441 AenGIyHisIleAenGIyGluAenTySerTyArgGIySerValIyRgIyCysAen 2460
 Db 7132 AACGACACATCAATGGGGAGAACTACAGTACCGGGCAGTGTGTACCAATGCAAT 7191
 Qy 2461 AlaGIyPheArgLeuIleGIyMetSerValArgIleCysGIyngInaSPH:HisTrpSer 2480
 Db 7192 GCTGGCTTCGCTGTATGGCATGTCTGTGGCATCTCCAGAGATCATCACTGGTGC 7251
 Qy 2481 GIyTyRThrProPheCysValProIleThrCysGIyHisProGIyAenProValAenGIy 2500
 Db 7252 GGCAAGACCCCTTCTGTCA---ATTACCTGTGGACACCGAGCAACCTGTCAACGGC 7307
 Qy 2501 LeuThrGIyngInaenGIyPheAenLeuAenAenValIyLysPheValCysAenProGIy 2520
 Db 7308 CTCACTCAGGGTAAACCACTTTAACTCAACGATGTGCTCAAGTTGTGTGCAACCTCGG 7367
 Qy 2521 TyrMetAlaGIyGluIyAlaAlaArgSerGIyngInaSerGIyGlnTrpSerAspMet 2540
 Db 7368 TATATGGTGAAGGGGCTGCTAGTCCCAATGCTGGCCAGCGGCAATGAGATGACATG 7427
 Qy 2541 LeuProThrCysArgIleIleAenCysThrAspProGIyHisGlnGluAenSerValArg 2560
 Db 7428 CTGCCCACTGCAGAAATCATCACTGACAGTCTGTGACACCAAGAAATATGTTGCT 7487
 Qy 2561 GlnValHisAlaSerGIyProHisArgPheSerPheGIyThrValIleSerTyRArgCys 2580
 Db 7488 CAGGTCCACGGCAGCGGCCGCCACAGTTCAGCTTCGGCACACACTGTCTTACCGGGTCC 7547
 Qy 2581 AenHisGIyPheTyRLeuLeuGIyThrProValLeuSerCysGlnGIyAenGIyThrTrp 2600
 Db 7548 AACCAAGGCTTCTACTCTCTGGGCAACCCAGTGTCTGACGTCCAGGGAGATGGCATGG 7607
 Qy 2601 AenArgProArgProGIyngInaCysLeuLeu-----ValSerCysGIyHisProGIySer 2617
 Db 7608 GACCGTCCCGGCCCGCCAGTGTCT--CTGAAGTATGAGTGTCTGTGGCCATCCGGGCTCC 7666
 Qy 2618 ProProHisSerGIyngInaMetSerGIyAenSerTyRThrValIyAlaValIyArgTySer 2637
 Db 7667 CCGGCTCACTCCAGATGTCTGGAGACAGTTACTGTGGAGGACGTGTGTCCGGTACAGC 7726
 Qy 2638 CysIleGIyLysArgThrLeuValGIyAenSerThrArgMetCysGIyLeuAenGIyHis 2657
 Db 7727 TGCATCGCAAGCGACTCTGGTGGAAACAGCACCCGATGTGGGCTGATGAGAAC 7786
 Qy 2658 TrpThrGIySerLeuProHisCysSerGIyThrSerValIyGIyValCysGIyAenProGIy 2677
 Db 7787 TGAATGGCTCCCTCCCTCACTGTGCAGGAACAGCGTGGAGATTGGCGTGAACCTGG 7846
 Qy 2678 IleProAlaHisGIyLysArgLeuGIyAenSerPheAenProGIyThrValMetArgPhe 2697
 Db 7847 ATCCCGGCTCATGTGATCGTTGGGGGACAGCTTTGATCCAGGACACTGTATGCGCTTC 7906
 Qy 2698 SerCysGIyAlaGIyHisValLeuArgGIySerSerGIyArgThrCysGlnAlaAenGIy 2717
 Db 7907 AGCTGTGAAGCTGGCCAGCTGTCCGGGGATGTGAGAGCGCACTGTCAAGCCAAATGGC 7966
 Qy 2718 SerTrpSerGIySerGIyngInaProGIyCysGIyValIleSerCysGIyAenProGIyThrPro 2737

Db 7967 TCGTGAAGCGGCTCGAGCCTGAGTGTGAGTATCTTTGGGAAACCTTGGACATCCA 8026
 Qy 2738 SerAenAlaArgValIlePheSerAenGIyLeuValIlePheSerSerIleValTyRGIy 2757
 Db 8027 AGTATGCCCCAGATTGTGTATGATGGCCGTGGTTTCTCAGCTCATGTCTATGATAG 8086
 Qy 2758 CysArgGIyGluIyTyRValAlaThrGIyLeuLeuSerArgHisCysSerValAenGIyThr 2777
 Db 8087 TGGCGGAGAGATCTAGCCACAGGCTGTGACCGGTCACTCTCGGTCAATGTATCC 8146
 Qy 2778 TrpThrGIySerAenProGIyCysLeuValIleAenCysGIyAenProGIyIleProAla 2797
 Db 8147 TGCACAGGACGTACCTGTAGTGCCTGTCTTAATCTGTGGTGAACCTGGGATTCAGGC 8206
 Qy 2798 AenGIyLeuArgLeuGIyAenAenPheArgTyRLeuSerValIleThrTyRGIyCysVal 2817
 Db 8207 AATGGCTTCGGCTGGGCAATGACTTCAAGTACAAACATGTGACATACATGTGTCTC 8266
 Qy 2818 ProGIyTyRMetMetGluSerHisArgValSerValLeuSerCysThrLysAspArgThr 2837
 Db 8267 CTTGGCTATATGATGAGTGCATATAGATATCTGTCTGAGCTGCACCAAGACCGAGCA 8326
 Qy 2838 TrpAenGIyThrLysProValCysLysAlaLeuMetCysLysProProProLeuIlePro 2857
 Db 8327 TGAATGGAACCAAGCCGCTGTGCAAGCTCTCATGTGCAAGCCACCTCCGCTCATCCC 8386
 Qy 2858 AenGIyLysValIleGIySerAenPheMetTrpGIySerSerValIleThrTyRAlaCysLeu 2877
 Db 8387 AATGGAAAGGTGGGGGTGTGACTTCAATGGGGGCTCAAGTGTGACTTATGACTGTCTG 8446
 Qy 2878 GlnGIyTyRngInaLeuSerLeuProAlaValPheThrCysGIyGluIyAenGIySerTrpThr 2897
 Db 8447 GAGGGATACCACTCTCCCTCCCGCGGTGTTCACCTGTAGGGAATGGTCTTGACCC 8506
 Qy 2898 GIyGlnLeuProGIyngInaPheProValPheCysGIyAenProGIyValIleProSerArgGIy 2917
 Db 8507 GGAAGAGCTGCCTCAGTGTTCCTGTGTTCGCGGGATCCTGTGTCCCGTCCGTTGGG 8566
 Qy 2918 ArgArgGluAenPheArgGIyPheSerTyRArgSerSerValSerPheSerCysHisProPro 2937
 Db 8567 AGGAGAAAGACCGAGGCTTCTCTACAGGTCATCTGTCTCTCTCGGCATCCCTCT 8626
 Qy 2938 LeuValLeuValGIySerProArgArgPheCysGlnSerAenArgGIyThrTrpSerGIyThr 2957
 Db 8627 CTGGTGTGTGGGCTCTTCCAGCGAGTGTTCAGTCAAGTATGAGATGAGTGGCACCC 8686
 Qy 2958 GlnProSerCysIleAenProThrLeuThrThrCysAlaAenProGIyValIleProGlnPhe 2977
 Db 8687 CAGCCAGCTGCATATGATCCGACCTGACACCGTGTGCGGACCTGTGTGTCCACAGTTT 8746
 Qy 2978 GlnIleGlnAenAenSerGIyngInaGlnValGIySerThrValIlePheArgCysGln 2997
 Db 8747 GGGATACAAACAATTTCCAGGGCTTACAGGTTGGAAGCACAGTCTTCCGTTCGTTGCA 8806
 Qy 2998 LysGIyTyRLeuLeuGIyngInaSerThrThrArgThrCysLeuProAenLeuThrTrpSer 3017
 Db 8807 AAAGGCTACTGTCTTCAGGGCTCCACACAGGACCTGTCTCCAAACCTGACTGGAGT 8866
 Qy 3018 GlnThrProProAenCysValProHisHisCysArgGlnProGIyThrProThrHisAla 3037
 Db 8867 GGAACCCCACTGACTGTGTCCCAACACAGGACCGAGGACGAGACCCCAACCAATGCC 8926
 Qy 3038 AenValGIyAlaLeuAenPheProSerMetGIyTyRThrLeuIleThrProAlaArgArg 3057
 Db 8927 AACGTGGGGGCTCGAATTTGCTCTCATAGGCTACACGCTCATTTCTCGCAGAGG 8986
 Qy 3058 AlaSerProSerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGIyGln 3077
 Db 8987 GCTTCTCTCTCAAGGGTGGCTCCAGACACCGCACTCTCAAGCGGATGGCACTGGACAG 9046
 Qy 3078 AlaSerArgProSerAlaIleArgSerGIyProValGIyAenProSerThrLeuProGIy 3097
 Db 9047 GCAAGCGGCCCATGTCTGTGAGGTCCGGCCAGTGGGAGACCATCAACACTGCCCGGG 9106

QY 3098 SerHisArgSerProIyAsp 3104
 DB 9107 AGCCACCGCTCACCCAGAGCT 9127
 RESULT 6
 ID ADH71135 standard; DNA; 10655 BP.
 XX ADH71135;
 AC ADH71135;
 DT 25-MAR-2004 (first entry)
 XX
 DE Human gene of the invention NOVA4 SEQ ID NO:31.
 XX
 KW ds; gene; human; cytostatic; immunomodulator; neuroprotective; nootropic;
 KW anorectic; antidiabetic; antimicrobial; antilipaeitic; gene therapy;
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
 KW obesity; diabetes; infectious disease; metabolic syndrome X;
 KW dyslipidaemia.
 XX
 OS Homo sapiens.
 PN WO2003102155-A2.
 PD 11-DEC-2003.
 XX
 PF 03-JUN-2003; 2003WO-US017430.
 XX
 PR 03-JUN-2002; 2002US-0385120P.
 PR 04-JUN-2002; 2002US-0385784P.
 PR 05-JUN-2002; 2002US-0386041P.
 PR 05-JUN-2002; 2002US-0386047P.
 PR 06-JUN-2002; 2002US-0386376P.
 PR 06-JUN-2002; 2002US-0386453P.
 PR 06-JUN-2002; 2002US-0386864P.
 PR 07-JUN-2002; 2002US-0387016P.
 PR 07-JUN-2002; 2002US-0387969P.
 PR 07-JUN-2002; 2002US-0388161P.
 PR 07-JUN-2002; 2002US-0388931P.
 PR 07-JUN-2002; 2002US-0386942P.
 PR 07-JUN-2002; 2002US-0386971P.
 PR 07-JUN-2002; 2002US-0387262P.
 PR 08-JUN-2002; 2002US-0296960P.
 PR 10-JUN-2002; 2002US-0387400P.
 PR 10-JUN-2002; 2002US-0387535P.
 PR 11-JUN-2002; 2002US-0387610P.
 PR 11-JUN-2002; 2002US-0387625P.
 PR 11-JUN-2002; 2002US-0387634P.
 PR 11-JUN-2002; 2002US-0387668P.
 PR 11-JUN-2002; 2002US-0387696P.
 PR 11-JUN-2002; 2002US-0387702P.
 PR 11-JUN-2002; 2002US-0387836P.
 PR 11-JUN-2002; 2002US-0387859P.
 PR 12-JUN-2002; 2002US-0387933P.
 PR 12-JUN-2002; 2002US-0387934P.
 PR 12-JUN-2002; 2002US-0387960P.
 PR 12-JUN-2002; 2002US-0388022P.
 PR 12-JUN-2002; 2002US-0388096P.
 PR 13-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389118P.
 PR 14-JUN-2002; 2002US-0389120P.
 PR 14-JUN-2002; 2002US-0389144P.
 PR 14-JUN-2002; 2002US-0389146P.
 PR 17-JUN-2002; 2002US-0389729P.
 PR 17-JUN-2002; 2002US-0389742P.
 PR 18-JUN-2002; 2002US-0389844P.
 PR 19-JUN-2002; 2002US-0390066P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 21-JUN-2002; 2002US-0390763P.
 PR 17-JUL-2002; 2002US-0396706P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402156P.

PR 09-AUG-2002; 2002US-0402256P.
 PR 09-AUG-2002; 2002US-0402389P.
 PR 12-AUG-2002; 2002US-0402786P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402819P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 12-AUG-2002; 2002US-0402832P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403531P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 13-AUG-2002; 2002US-0403617P.
 PR 15-AUG-2002; 2002US-0403617P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-0412731P.
 PR 30-SEP-2002; 2002US-0414801P.
 PR 30-SEP-2002; 2002US-0414839P.
 PR 30-SEP-2002; 2002US-0414840P.
 PR 30-SEP-2002; 2002US-0414954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-00423798.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Alsbrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,
 PI Eutenberg S, Gangolli EA, Gerlach VL, Gorman L, Gunther E, Guo X,
 PI Gusev VY, Herrmann JU, Ji W, Kekuda R, Li L, Liu X, Macdougall JR,
 PI Maciachian T, Malynkar UM, Mezick AJ, Miller I, Mishra VS;
 PI Padigaru M, Paturajan M, Pena CE, Peyman JA, Raha D, Raetzelli L,
 PI Rieger DK, Rothenberg ME, Scioe P, Shenoy SG, Shinkets RA;
 PI Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M,
 PI Zhong H;
 XX
 DR WPI; 2004-081935/08.
 DR P-Psdb; ADH71136.
 XX
 PT New NOVA polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVA-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 4; SEQ ID NO 31; 1880pp; English.
 XX

CC The invention relates to a novel isolated polypeptide (NOVA). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antilipaeitic activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVA polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing. The
 CC treating or preventing NOVA-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemia. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC encodes a NOVA polypeptide of the invention.
 XX

Sequence 10655 BP; 2255 A; 3236 C; 2879 G; 2285 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	10655
Score:	16319.00	Matches:	3009
Percent Similarity:	97.40%	Conservative:	21
Best Local Similarity:	96.72%	Mismatches:	45
Query Match:	96.08%	Indels:	37
DB:	12	Gaps:	10

US-10-016-248-2 (1-3104) x ADH71135 (1-10655)

QY	2	AlaGlyAlaProProProAlaLeuLeuLeuProCysSerLeuIle---SerAspCys---	19
Db	731	TCGGTTCAACGGAGCCAGCCTCCAGCCCCCGTTATACGACAGCAAGACTGGCTGCCAAC	790
QY	20	CysAlaSerAaenGlnArgHisSerValGlyValGly-----ProSerGluVal	36
Db	791	TGCACCTTCACATCGGATGGTGCACCAACCGGACGGCGGATTCAGTCCCAATTACCA--GTC	849
QY	37	LybelyegInIleGluLeuLysSerArgGlyValLybLeuMetProSerLybAspAsnSer	56
Db	850	AAGAAGCAAAATTGAATTGAAGTCTCCAGAGTGTGAAGCTGAATGCCACGAAAGACAAACAGC	909
QY	57	GlnlybThrSerValLeuThrGlnValGlyValSerGlnGlyHisAspMetCysProAsp	76
Db	910	CAGAAAGAGCTGTGGTGAATCTCAGGTTGTGTGTCTCCAGACATAAATATGTCTCAAC	969
QY	77	ProGlyIleProGluArgGlyLybArgLeuGlySerAspPheArgLeuGlySerSerVal	96
Db	970	CCTGGCATACCCGAAAGGGGCAAAAGATCAGGCTCGGATTTCAAGTTAGATTCACAGCTC	1029
QY	97	GlnPheThrCysAaenGluGlyTybAspLeuGlnGlySerLybArgIleThrCysMetLys	116
Db	1030	CAGTTCACCTTCACAGAGGGGCTATGACCTGCAGAGGCTCCAGCGGATACCTGTAAGAA	1089
QY	117	ValSerAspMetPheAlaAlaTrpSerAspHisArgProValCysArgAlaArgMetCys	136
Db	1090	GTGAGCGACATGTTTGGCGCGCTGGAGCGACACAGGCCAGCTGCTGCCAGCCCGCATGT	1149
QY	137	AspAlaHisLeuArgGlyProSerGlyIleIleThrSerProAsnPheProIleGlnTr	156
Db	1150	GATGCCCAACCTTCAGAGGCCCTCGGGCATCATCACTCCCAATTTCCCATTCACGAT	1209
QY	157	AspAsnAsnAlaHisCysValTrpIleIleThrAlaLeuAsnProSerLybValIleLys	176
Db	1210	GACAAACAATGACACACTGTGTGTGATCATCAAGACACTCAACCCCTCCAAAGTAAATCAAG	1268
QY	177	LeuAlaPheGluGluPheAspLeuGluArgGlyTybAspThrLeuThrValGlyAspGly	196
Db	1270	CTCCGCTTTGAGGAGTTGATTGTGGAGAGGGCTATGACACCTGACGCGTCCGTGATGT	1329
QY	197	GlybLysPheGlyAspGlnLysThrValLeuLysTyrMetSerGlnAsnAlaCysSerAspSer	216
Db	1330	GGTACAGATGGGAGCCAGAAAGACAGTTCTCAACG-----	1365
QY	217	ProHisThrProGlySerArgIleProGluSerMetSerGlyAspIleTrpArgGlnLys	236
Db	1366	-----CTGACAGGATACATCGGTCGCCGATCTCATTTGCACACCAAT---CATCAATG	1416
QY	237	TyrThrValLeuGluIleCysArgAspIleSerSerSerAspAlaArgSerGlySerVal	256
Db	1417	TGGCTCTCTCTCCAG-----ACTGATGGC--AGTGGCAGT---	1449
QY	257	ArgLysSerProLysThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlu	276
Db	1450	-----TCCCTGGGATTCAAAGCTTCTTATGAAGAAGTCCAG	1485
QY	277	GlnGlySerCysGlyAspProGlyIleProAlaTybArgArgGlnGlySerArgPhe	296
Db	1486	CAGGGCAGATTCCGGGACCTCTGCATACCTGCATATGGCCGAGGAAAGCTCCCGGTTT	1545
QY	297	HisIleGlyAspThrLeuLysPheGluCysGlnProAlaPheGluLeuValGlyGlnLys	316
Db	1546	CGCCACGGTGCACACTCAATTTGATGAGCCAGCCCGCTTTGAGCTGGTGGACAGAG	1605

QY	317	AlaIleThrCysGlnIlyAsnAsnGlnTrpSerAlaIlyGlyProGlyCysValPheSer	336
Db	1606	GCAAATCACATGCGAAAGAAATACCAATAGTGTGGCTAAGAAACAGGCGCTGTGTTCC	1663
QY	337	CysPhePheAsnPheThrSerProSerGlyValValIleUserProAsnTyrProGluAsp	356
Db	1666	TGCTTCTTCACTTACCAAGCCCGCTGGGGTGTGCTGTCTCCCACTACCAAGAGAC	1722
QY	357	TyrGlyAsnHisIleLeuHisCysValTrpLeuIleLeuAlaArgProGluSerArgIleHis	376
Db	1726	TATGGCAACCACTCCACATGTGTGTGGCTCATCTGGCAGGCGCTGAGGCGGCATCCAC	1789
QY	377	LeuAlaIlePheAsnAspIleAspValGluProGlnPheAsnPheLeuValIleIlyAspGly	396
Db	1786	CTGGCGCTTCAACGACATATGACCTGTGAGGCTCAGTTTGTATTTCTGTGATCAAGATGGG	1844
QY	397	AlaTrpAlaGluAlaProValIleuGlyThrPheSerGlyAsnGlnLeuProSerSerIle	416
Db	1846	GCCACCGCGGAGGCGCCCGCTCTGGGACACTTCTAGGAACCACTTCCCTTCCATC	1907
QY	417	ThrSerSerGlyHisValAlaArgLeuGlnPheGlnTrpAspHisSerThrGlyIlyAspArg	436
Db	1906	ACAACACAGTGGCCAGCTGGCCCGCTCTCGAGTTCCAGACTGACCACTCCACAGGAAAGG	1966
QY	437	GlyPheAsnIleThrPheThrThrPheArgHisAsnGlyCysProAspProGlyValPro	456
Db	1966	GGCTTCAACATCACCTTTTATCCAGTAGTGCCTTCCAAACGATGCGCCGAGATCCGCGCTTCCA	2022
QY	457	ValAsnGlyIlyValArgPheGlyAspSerIleuGlnIlySerSerIleSerPheIleuGlyCys	476
Db	2026	GTAATATGGCAACGGATTGGGGACAGCTCCAGAGTGGGAGCTCATCTCTCTCTGT	2089
QY	477	AspGlnGlyPheLeuGlyThrGlnIlySerGluThrIleThrCysValIleuIlySerGluGly	496
Db	2086	GATGAAGGCTTCTTGGGACTCAGGGCTCAGAGACATCACTCGTGCTCTGAAGAGAGGC	2144
QY	497	SerValValTrpAsnSerAlaValIleuArgCysGluAlaProCysGlyIlyHisIleuThr	516
Db	2146	AGCGGTGCTGTGGAACAGCGCTGTGCGGTGTGAAAGCTCCCTGTGTGTGTACCTGACT	2205
QY	517	SerProSerGlyIlyThrIleLeuSerProGlyTrpProGlyIlyPheTyrIlyAspAlaIleuSer	536
Db	2206	TGCGCCACCGGACCATCTCTCTCGGAGCTGGCTGCTTCTCAACAGAAAGCTTGGAGC	2265
QY	537	CysAlaTrpValIleGluAlaGlnProGlyTyrProIleIlyHisIleThrPheAspArgPhe	556
Db	2266	TGTGCTGTGGTGATTGAGGCCAGCCAGGACTACCCCATCAAAATCACTTGACACAGGTTCC	2322
QY	557	LysThrGluValaIleuTyrAspThrIleuGlnValaArgAspGlyArgThrTyrSerAlaPro	576
Db	2326	AAAAACCGAGGTCAACTATGACACCCCTGGAAAGTACCGCATGTGGAGCTTATCTACAGCGCCC	2389
QY	577	LeuIleGlyValaTyrHisGlyIlyThrGlnValaProGlnPheLeuIleSerThrSerAsnTyr	596
Db	2386	TTGATCTGGGGTTTACACAGGAGCCAGGATTTCCCACTTCAATGACACAGCAACTAC	2444
QY	597	LeuTyrIleuLeuPheSerThrAspIlySerHisSerAspIleGlyPheGlnIleuArgTyr	616
Db	2446	CTTACCTCTCTTCTCTTACCGGACAAAGTAGTACTCGGACATCCGGCTTCCAGCTCCGCTAT	2509
QY	617	GluThrIleThrLeuGlnSerAspHisCysIleuAspProGlyIleProValaIleuGln	636
Db	2506	GAGGCTATAACCTGACAGTACAGACCACTGTCTGGATCCAGGAATCCAGTAATATGACAG	2565
QY	637	ArgHisGlyAsnAspPheTyrValGlyAlaIleuValThrPheSerCysAspSerGlyTyr	656
Db	2566	CGTACATGGAAATGACTTCTTACGTGGCGCGCTGTGGACTTCACACTGTGATCTGGCGTAC	2625
QY	657	ThrLeuSerAspGlyGluProIleuGlnCysGluProAsnPheGlnTrpSerArgAlaIleu	676
Db	2626	ACATTAAATGACGGGGAGCTCTGTGAGTGTAGGCCCAATCTTCCAGTGAAGCCGGGCTTG	2685
QY	677	ProSerCysGluAlaIleuCysGlyGlyPheIleGlnIlySerSerGlyIlyThrIleuSer	696

```
Db 2686 CCCAGTTGGAAGCTCTGTGTGCTGCTTCAATCAAGCTCCAGTGGGACCATCTTGTG 2745
Qy 697 ProGlyPheProAspPheTyrProAsnAsnLeuAsnCystrProIleIleGlyThrSer 716
Db 2746 CCAAGTTTCCCTGACTTCTTACCCCAACAATTGAATCAGCTGATTAATGAAACATCT 2805
Qy 717 HisGlyLysGlyValPhePheThrPheHisThrPheHisLeuGlySerGlyHisAspTyr 736
Db 2806 CATGGCAAGGGGTGTCTTCTTCACTTCCACACCTTCCACTGGAAAGTGGCCATGACTAC 2865
Qy 737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db 2866 CTCCCATCACTGAGAACGCGAGCTTCCAGCCCTGAGGAGCGTAACTGATCTCGG 2925
Qy 757 LeuProAlaProIleSerAlaGlyLeuThrGlyLysAsnThrAlaGlnValArgPheIle 776
Db 2926 CTGCGAGCTCCCATCAAGCGCTGGGCTGTATGGCAACTTCACTGCGCCAGCTTCATC 2985
Qy 777 SerAspPheSerMetSerTyrGlyGlyPheAsnIleThrPheSerGlyTyrAspLeuGlu 796
Db 2986 TCTGATTTCTCCATGTCTATGAAGATTCAAATCACTTCTTCAAGTACGACTTGGAG 3045
Qy 797 ProGlyGluGluProGlyValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyVal 816
Db 3046 CCGTGTGAGGAGCCCGAGGCTCCAGCTACAGCATCCGGAAGGGGCTTGAGTTGGGCTG 3105
Qy 817 GlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeuGlnGlyThrAlaArgIle 836
Db 3106 GGGACACCTTGAACCTTCTCTGCTTCCCGGGATCCGCTCGAGGGGACCGCCGCTATC 3165
Qy 837 ThrCysLeuGlyGlyValArgArgLeuTyrPheSerSerProLeuProArgCysValAlaGlu 856
Db 3166 ACGGCTGTGGGGGCAACGCGCTGTGGAGCTCGCTTGCACAGTGTGTGCTGAG 3225
Qy 857 CysGlyAsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsn 876
Db 3226 TGTGGGAATTCAGTCAAGGACCTCAAGGATCACTTGTGTGCTTCCCACTTCTGTGAAC 3285
Qy 877 TyrAsnAsnAsnHisGlyCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGln 896
Db 3286 TACAAATACATCATGATGATGATCTACTCTCAATCCAGCCAGGAGGAAAGGAAATTCAG 3345
Qy 897 LeuLysAlaIArgAlaPheGlyLeuLeuSerGlyGlyAspValIleuValTyrAspGlyAsn 916
Db 3346 CTAAAGCCAGGGGCAATTCGAACCTCCGAGAGAGATGCTCTCAAGGTTTATGATGGCAAC 3405
Qy 917 AsnAsnSerAlaArgLeuLeuGlyValPheSerHisSerGluMetSerGlyValThrLeu 936
Db 3406 AACACATCCGCCCCGTTTCTGCGGAGTTTAAAGCATTTCTGAGATGATGGGGTGACTTGG 3465
Qy 937 AsnSerThrSerSerSerLeuTyrPheAspPheIleThrAspAlaGluAsnThrSerLys 956
Db 3466 AACAGACATCCAGGAGCTGTGTGCTTGAATTCATCATCATGATCGTGAAGAACCCAGCAAG 3525
Qy 957 GlyPheGluLeuHisPheSerSerPheGlyLeuLeuIleLysCysGlyLysAspProGlyThrPro 976
Db 3526 GGGTTGAACCTGCACTTTCAGATTTCAGTTCATCACTCAAAATGTGAGAGCCACAGAAACCCCC 3585
Qy 977 LysPheGlyTyrLysValHisAspGlyGlyHisPheAlaGlySerSerValSerPheSer 996
Db 3586 AAGTTTGGCTACAAAGTTTCAATGATGAGTCAATTTTGGAGGAGCTCCGTGCTTCTTCAAC 3645
Qy 997 CysAspProGlyTyrSerLeuArgLysSerGlyLeuLeuCysLeuSerGlyGluArg 1016
Db 3646 TGTGACCTCGATACAGCTTCGCGGGTATGTAGAGAGCTGTGTGCTGATGAGAGCGC 3705
Qy 1017 ArgThrTyrAspArgProLeuProThrCysValAlaGlyCysGlyGlyThrValArgGly 1036
Db 3706 CCGACTGGGACCGGGCTTCTGCGCACCTGTGTGCGGAGTGTGAGGAGCAGTGTAGAGA 3765
Qy 1037 GluValSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsn 1056
Db 3766 GAGGTGTGGGCGAGGCTGTCTACCCGGGATTCAGCTCCCTATGAAACAATCTCAAC 3825
Qy 1057 CysIleTyrThrIleGlnAlaGlyCysThrIleGlyLeuHisPheLeuValPhe 1076
Db 3826 TGCATCTGACCATTCAGAACAGAGCGGCTGACCATTTGGGCTTACACTTCTGAGTGT 3885
Qy 1077 AspThrGluGluValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyVal 1096
Db 3886 GACACAGAGAGAGGTTCACACAGCTGTGCGCATCTGGGATGGGCTCTGTGAGAGCGGGT 3945
Qy 1097 LeuLeuLysGlyLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSer 1116
Db 3946 CTGCTAAAGAGCTGATGTGGCCCGGCTTGGCCAAAGACCTGCATAGCACTTCAACTCG 4005
Qy 1117 ValValLeuGlnPheSerThrAspPheThrSerLysGlnGlyPheAlaIleGlnPhe 1136
Db 4006 GTGCTCTCAGATTGAGCATGACTTCTTACACAGCAAGCAGGCTTTCCTCAATTT 4065
Qy 1137 SerValSerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSer 1156
Db 4066 TCAAGTTCCACAGCAACGCTTGCATATGACCTGGGATCCGCAAAATGGAGTGGAGT 4125
Qy 1157 GlyAspSerTyrGlyAlaGlyAspSerThrValPheGlnCysAspProGlyTyrThrAlaLeu 1176
Db 4126 GGTGACAGTTGGGAGAGCCGGGCACTCCACAGTGTTCAGTGTGACCTTGGCTACGCGCTG 4185
Qy 1177 GlnGlySerAlaGluIleSerCysValLysIleGluAsnArgPhePheTyrGlnProSer 1196
Db 4186 CAGGAGAGTGCAGAGATCAGCTGTGTGAAGATCGAGAACAGATTCTTCTGCGAGCCAGC 4245
Qy 1197 ProProThrCysIleAlaProCysGlyGlyAsnLeuThrGlyProSerGlyValIleLeu 1216
Db 4246 CCGCAACATGATGCTCTCTGCGGGGAGAACCTGCAGAGAACATCTGGAGTCACTTC 4305
Qy 1217 SerProAsnTyrProGluProTyrProProGlyLysGlyCysAspTyrPheValThrVal 1236
Db 4306 TCACCAATTAACCGAACCCTTACCCGCAAGCAAGAGTGTGACTGGAAGTACCTC 4365
Qy 1237 SerProAspTyrValIleAlaLeuVal---PheAsnIlePheAsnLeuGlnProGlyTyr 1255
Db 4366 TCACAGACTGACGTCATCGCCCTGTGTATTTCTCCCAAGCTTAACTGCGAGCTGCGTAT 4425
Qy 1256 AspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyr 1275
Db 4426 GACTTCTCCATATCTAGACAGAGCGGAGCTTCTAGCCCTTCATAGGAAGGCTTCTAT 4485
Qy 1276 GlySerGlnLeuProGlyArgIleGlySerSerSerAsnSerLeuPheLeuAlaPheArg 1295
Db 4486 GGGTCCAGCTCCAGGCGCGCATTTGAAGCAGCAGCAAGCCCTTCTCGCTCGC 4545
Qy 1296 SerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlu 1315
Db 4546 AGCATGCACTGTGACAAATGCTGCTGTGTATGATCACTTCCAGAAACCCCGGGAG 4605
Qy 1316 SerCysPheAspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeu 1335
Db 4606 TCATGTTTGTGCTGTGTTCCATCAAGAAAGCAGCAGCGGTGGGCTCGACCTGAACTG 4665
Qy 1336 GlySerSerValThrTyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeu 1355
Db 4666 GGGCTCTCCGTACCTACTCTGCAACGGGGGCTACGAAAGTTGAGGAGCACTCGACCTG 4725
Qy 1356 SerCysIleLeuGlyProAspGlyLysProValTyrAsnAsnProArgProValCysThr 1375
Db 4726 AGCTGCACTCTGGGGCTGATGTGGAAGCCGCTGTGGAACAATCCCGGCGAGTCTCAACA 4785
Qy 1376 AlaProCysGlyGlyGlnTyrValGlySerAspGlyValValLeuSerProAsnTyrPro 1395
Db 4786 GCCCCTGTGGGAGACAGTATGTGGTTCGAGAGAGTGTCTTGTCCCCCAACTACCC 4845
Qy 1396 GlnAsnTyrThrSerGlyGlnIleCysLeuTyrPheValThrValProLysAspTyr--- 1414
Db 4846 CAGAACTACACAGTGAAGATCTGCTTGTATTTTGTATCTGTGCCAAAGGACTATGGT 4905
```

```
QY 1415 ValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspValValGluValHis 1434
DB 4906 GTGGTGTGGCCAGCTTGCCTTCTTTCACAGCGCCCTCAACAGAGTGTGAGAGTTTAC 4965
QY 1435 AspGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGlu 1454
DB 4966 GACGGCCACAGCCAGCACTCGCGGTCTCTCAGCTCCCTCGGGGTCCATACAGGTGA 5025
QY 1455 SerLeuProLeuAlaThrSerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAla 1474
DB 5026 TCAGTGCCTTGGCCAGCTCCAAATCAAGTTCAATTAGTTCACGCGCAAGGCGCTCGCA 5085
QY 1475 ProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSerAlaThrGlyCys 1494
DB 5086 CCAAGCCAGAGGCTTCCACTTGTCTTCTCAACAGCGGTTCTTCCAGCCAGCGCAGAGTGC 5145
QY 1495 SerSerValProGluProArgTyrGlyLysArgLeuGlySerAspPheSerValGlyAla 1514
DB 5146 AGCTGTGGCCGGAACCCCGCTATGGCAAGGCGGTGAGTGACTTCTCGGTGGGGCC 5205
QY 1515 IleValArgPheGlyCysAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGlyCys 1534
DB 5206 ATCTTCGCTTCAATGCACTCCGCGCTATGCTCCGCAAGGCGTCCAGAGATGAGTGC 5265
QY 1535 LeuProValProGlyAlaLeuAlaGlnThrAsnValSerAlaProThrCysValValPro 1554
DB 5266 CTCCCTGTGCTGGGGCTTGGCCCAAGAGATGTCTCAGCGCCAGCGTGTGTGGCCG 5325
QY 1555 CysGlyGlyAsnLeuThrGluArgArgGlyThrIleLeuSerProGlyPheProGluPro 1574
DB 5326 TGTGAGGCAACCTCAAGAGCGAGGCGCACATCCGTCCCTGCTCCAGAGCG 5385
QY 1575 TyrLeuAsnSerLeuAsnCysValTTrpLysIleValValProGluGlyAlaGlyIleGln 1594
DB 5386 TACTTCAACAGCTCAACTGTGTGTGAGATCGTGTGCCCGAAGGCGGTGCATCAG 5445
QY 1595 IleGlnValValSerPheValThrGluGlnAsnTrpAspSerLeuGluValPheAspGly 1614
DB 5446 ATCCAGTGTGCACTTGTGTACAGAGCAAGACGTGGAGCTCGTGAAGATTTGATGT 5505
QY 1615 AlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeu 1634
DB 5506 GCATATAACCTGTAACCATCTGGGAGTTTCTCAGAACACCGTGCCTCTCTG 5565
QY 1635 AsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSerValSerAlaAla 1654
DB 5566 AACAGCACTTCAACCAAGCTTACCTTCACTTCTACATATCAGATATCAGGTATCTGAGCT 5625
QY 1655 GlyPheHisLeuGlnTyrTyrLeuThrValGlyLeuSerSerCysProGluProAlaValPro 1674
DB 5626 GGCTTCCACTTGGAGTACAAACGAGTGGCTGAGCAGTTGTGCGAACCCTGCTGTGCC 5685
QY 1675 SerAsnGlyValLysThrGlyGluArgTyrLeuValAsnAspValValSerPheGlnCys 1694
DB 5686 AGTAACGGGGTGAAGACTGGGAGCGCTACTTGGTGAATGATGTGTCTTTCAGTGT 5745
QY 1695 GluProGlyTyrAlaLeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArg 1714
DB 5746 GAGCGGGGATGTGCTTCCAGGGCCAGCGCCACATCTCTGATGCCCGGAACAGTGGG 5805
QY 1715 ArgTrpAsnThrProProProLeuCysIleAlaGlnCysGlyGlyThrValGluGluMet 1734
DB 5806 CGATGGAACATACCTCTCTCCACTGTGATGGCAAGTGTGGGGAACAGTGAAGAAATG 5865
QY 1735 GluGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsnMetAspCysSer 1754
DB 5866 GAGGGGGTGAATCCGAGCCCGGCTTCCAGGCAACATACCCCAATTAACATGACTGCTCC 5925
QY 1755 TrpLysIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThr 1774
DB 5926 TGGAAATATAGCACTGCGCGGTGGCTTGGAGACTCAATCCAGTCTGAACTCTCCACC 5985
QY 1775 GluProAsnHisAspTyrIleGluIleArgAsnGlyProTyrGluThrSerArgMetMet 1794
DB 5986 GAGCCCAACCAAGCACTCATAGAAATCCGAAATGGCCCCCTATGAGACCAAGCCGATATG 6045
QY 1795 GlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSerHisGluThrThr 1814
DB 6046 GGAAGATTCAAGTGAACGAGCTTCCAAAGCTCCCTCTCTCAGCTCCCAAGAACCAACC 6105
QY 1815 ValTyrPheHisSerAspHisSerGlnAsnArgProGlyLysPheLysLeuGluTyrGlnAla 1834
DB 6106 GTGATATTCCACAGCCAGCACTCCCAAGATTCGCGCAGATTCAGCTGGAGTATCAGAGCC 6165
QY 1835 TyrGluLeuGlnGlyCysProAspProGluProPheAlaAsnGlyIleValArgGlyAla 1854
DB 6166 TATGAACCTTCAAGGTGCCCAAGCCAGAGCCCTTTGCCATGGCATTTGAGGGAGAGCT 6225
QY 1855 GlyTyrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTyrGlnLeuThrGly 1874
DB 6226 GGCTACAAAGTGGGACATCATGATCCTTGAAGTGCCTCCCGGGGTATCATTAATGACTGGC 6285
QY 1875 HisProValLeuThrCysGlnHisGlyThrAsnArgAsnTrpAspHisProLeuProLys 1894
DB 6286 CACCTGTCTCTCAAGTGTCAACATGGCACCAACCGGAACGTGGGACCAACCCCTGCCCCAG 6345
QY 1895 CysGluValProCysGlyGlyAsnIleThrSerSerAsnGlyThrValTyrSerProGly 1914
DB 6346 TGTGAATCCTTGTGGCGGGAAACATCATCTTCTTCAACGAGCATGTGTACTCCCGGGG 6405
QY 1915 PheProSerProTyrSerSerSerSerGlnAspCysValTTrpLeuIleThrValProIleGly 1934
DB 6406 TTCCCTAGCCCGATCTCAGCTCCAGAGCATGTGTCTTCAACGAGCATGTGTACTCCCGGGG 6465
QY 1935 HisGlyValArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThr 1954
DB 6466 CATGGCGTCCGCTCAACCTCAGCTGTCTCAGACAGAGCCCTCTGGAATTTTCATCACC 6525
QY 1955 IleTrpAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAla 1974
DB 6526 ATCTGGATGGGCCACAGCAACAGCAACGAGCTCGGCTCTTCAACCCGAGCATGGCC 6585
QY 1975 LysLysThrValGlnSerSerSerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAla 1994
DB 6586 AAGAAACAGTGAAGATTCATCAACCAAGTCTCTCAAGTTCCACCGATGAGCAGCC 6645
QY 1995 ThrGlyGlyIlePheAlaIleAlaPheSerAlaTyrProLeuThrLysCysProProPro 2014
DB 6646 ACAAGGGGGATCTTCCGCAATGACTTCTCCGCTTATCACTCAACCAATGCCCTCTGCC 6705
QY 2015 ThrIleLeuProAsnAlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIle 2034
DB 6706 ACCATCTCTCCCAACGCGCAAGTGTCAAGAGATGAAGAAATTCATATATGATGATC 6765
QY 2035 ValArgTyrArgCysLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLys 2054
DB 6766 GTACGCTACAGAGTGCCTCCCTGCTTAACTTATGGGGAATGAATTCGACTGCAAA 6825
QY 2055 LeuGlyThrTyrLeuGlnPheGluGlyProProProIleCysGluValHisCysProThr 2074
DB 6826 CTGGAACTTACCTGAGTGTGAAGGACCAACCCCGATATGTAAGTGAAGCACTGCCAACA 6885
QY 2075 AsnGluLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTyrProGlySerTyr 2094
DB 6886 AATGAGCTTGTGACAGCTCCACAGGGGTATCTTGAGCCAGAGCTACCTCGGAAGCTAT 6945
QY 2095 ProGlnPheGlnThrCysSerTTrpLeuValArgValGluProAspTyrAsnIleSerLeu 2114
DB 6946 CCCAGTTTCAAGACTGCTCTTGGCTGTGAGTGAAGGCCAGCATATAACATCTCCTC 7005
QY 2115 ThrValGluTyrPheLeuSerGlyLysGlnTyrAspGluPheGluIlePheAspGlyPro 2134
DB 7006 ACAGTGAAGTATCTTCCACAGAGAAAGCAATATGATGATTTTGTGATGTGCCA 7065
QY 2135 SerGlyGlnSerProLeuLeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleVal 2154
```



```
Db      7066 TCAAGACAGAGTCTCTGCTGANAAGCCCTCAGTGGAAATTACTCAGCTCCCTGATTGTC 7125
Qy      2155 ThrSerSerSerAsnSerValIyrrLeuAgtTrrSerSerAspHisAlaTyrrAsnArglys 2174
Db      7126 ACCAGCTCAAGCAACTCTGTGTACTGCTGCTGTGTATCATGTGATCAGCGCTTACATGGGAAAG 7185
Qy      2175 GlyPheIyLeIAsrGyrSerIaProTyrrCySerLeuProArgAlaProLeuHisGly 2194
Db      7186 GGGTTCAAGATCCGGCTATTTCAGCCCTTACTGACACCTGGCCAGGGCTTCACTCATGAGC 7245
Qy      2195 PheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPheGlyCyAsnAla 2214
Db      7246 TTCATCTTAGGCGCAGACGACGACCCAGCCGGGGCTCATCTGCTTGGCTGCAAGCC 7305
Qy      2215 GlyTyrrArgLeuValGlyHisSerMetAlaIleCyThrArgHisProGlnGlyTyrrHis 2234
Db      7306 GGGTACCGCTGGTGGGACACAGATGGCATCTGTACCCGGCACCCCGAGGCTACAC 7365
Qy      2235 LeuTrrSerGluAlaIleProLeuCySerGlnAlaLeuSerCySerGlyLeuProGlnAlaPro 2254
Db      7366 CTGTGAGAGCAAGCCATCTCTCTCTGTACAGCTCTTTCCTGTGGAGCTTCTCGAGGCCCC 7425
Qy      2255 LysAsnGlyMetValPheGlyLysGlyTyrrThrValGlyTyrrLysAlaValTyrrSerCys 2274
Db      7426 AAGAATGGAAATGGTGTGGCAGAGGATACACAGTGGGAACCAAGGCCATGTACAGCTGC 7485
Qy      2275 SerGluGlyTyrrHisLeuGlnAlaGlyAlaGlnAlaThrAlaGlnLysLeuAspThrGly 2294
Db      7486 AGTAAGGCTTACCACTTCCAGGCGAGGCGCTGAGGCCACTGAGAGTGTCTGGACACAGGC 7545
Qy      2295 LeuTrrSerAsnArgAsnValProProGlnCySerValProValThrCysProAspValSer 2314
Db      7546 CTATGGAGCAACCCCAATGTCTCCACACAGTGTCTCTGTGAGTCTTCTGTATGTACGT 7605
Qy      2315 SerIleSerValGlnHisGlyArgTrrArgLeuIlePheGlnThrGlnTyrrGlnPheGln 2334
Db      7606 AGCATACGCTGGAGCATGGCCGATGGAGGCTTATCTTTGAGACACAGTATCATGTTCCAG 7665
Qy      2335 AlaGlnLeuMetLeuIleCyAspProGlyTyrrTyrrTyrrThrGlnAlaArgValIleArg 2354
Db      7666 GCCCAGCTGATGCTCATCTGTGACCTGGCTACTACTACTGACCAAGGCTCATCTCCG 7725
Qy      2355 CySerGlnAlaAsnGlyLysTrrSerLeuGlyAspSerThrProThrCysArgIleIleSer 2374
Db      7726 TGTAGGCCCAATGGCAATGAGAGCTCGGGGACTTACGCCCACTGGCGAATCATCTCC 7785
Qy      2375 CySerGlyLeuLeuProIleProProAsnGlyHisArgIleGlyThrLeuSerValTyrrGly 2394
Db      7786 TGTGGAGAGCTCCCGATCTCCCGCATGGCCCATGGCCACCGCATGGAACTGTCTGTACCGG 7845
Qy      2395 AlaThrAlaIlePheSerCyAsnSerGlyTyrrThrLeuValGlySerArgValArgGln 2414
Db      7846 GCAACAGCCCATCTTCTCTGCAATTCGGATACACACTGTGGGCTCCAGGGTGGCTGAG 7905
Qy      2415 CyMetAlaAsnGlyLeuTrrSerGlySerGlyValArgCyLeuAlaGlyHisCysGly 2434
Db      7906 TGTATGGCCCAATGGCTCTGGAGTGGCTCTGAAAGTCCGCTGGCTGGAGACATCTGGG 7965
Qy      2435 ThrProGluProIleValAlaAsnGlyHisIleAsnGlyGluAsnTyrrSerTyrrArgIleSer 2454
Db      7966 ACTCCCTGAGCCCATGTGTCAACGGAACATCATGGGAGAACTACAGCTACCGGGGCGCT 8025
Qy      2455 ValValTyrrGlnCyAsnAlaGlyPheArgLeuIleGlyMetSerValArgIleCysGln 2474
Db      8026 GTGGTGTACCAATGCAATGCTGCTTCCCGCTGATCGGATGTCTGTGGCATCTGGCAG 8085
Qy      2475 GlnAspHisHisTrrSerGlyLysTyrrProPheCySerValProIleThrCysGlyHisPro 2494
Db      8086 CAGGATCATCATGCTGGGCAAGACCCCTTCTGTGTGCAATTAACCTGTGGACACCCA 8145
Qy      2495 GlyAsnProValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLys 2514

Db      8146 GGCNAACCTGTCAACGGCTCACTCAGGGGTAAACAGTTTAACCTCAACAGATGTGTCAG 8205
Qy      2515 PheValCyAsnAsnProGlyTyrrMetAlaGlnGlyAlaAlaArgSerGlnCysLeuAlaSer 2534
Db      8206 TTGTGTGGCAACCTTGGGTATATAGCTGAGGGGGCTGTAGGTCCAAATGCTGGCCAG 8265
Qy      2535 GlyGlnTrrSerAspMetLeuProThrCysArgIleIleAsnCySerThrAspProGlyHis 2554
Db      8266 GGGCAATGAGATGACATGTCTGGCCCACTGCAAGATATATCATGTAACATCTCGGACAC 8325
Qy      2555 GlnGluAsnSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyThr 2574
Db      8326 CAAGAAATAAGTGTGTGATGAGTCCAGCCGACAGGGCCGACAGGTTACGTTCCGACAC 8385
Qy      2575 ThrValSerTyrrArgCyAsnHisGlyPheTyrrLeuLeuGlyThrProValLeuSerCys 2594
Db      8386 ACTGTGTCTTACCGGTGCAACAGGCTTCTACTCTCTGGGACCCCACTGCTCAGCTGC 8445
Qy      2595 GlnGlyAspGlyThrTrrAspArgProArgProGlnCyLeuLeuValSerCysGlyHis 2614
Db      8446 CAGGAGATGGACATGGGACCGTCCCGCCCGCACTGTCTCTGTGTCTGTGCGCAT 8505
Qy      2615 ProGlySerProProHisSerGlnMetSerGlyAspSerTyrrThrValGlyAlaValVal 2634
Db      8506 CCGGCTCCCGCTCACTCCAGATGCTTGGAGACATTTATCTGTGGAGCAGTGGTG 8565
Qy      2635 ArgTyrrSerCysIleGlyLysArgThrLeuValGlyAsnSerThrArgMetCysGlyLeu 2654
Db      8566 CGGTACAGTGCATGGGCAAGGTACTCTGTGTGGGAACACAGCCCATGTGTGGCTG 8625
Qy      2655 AspGlyHisTrrThrGlySerLeuProHisCysSerGlyThrSerValGlyValCysGly 2674
Db      8626 GATGGAACCTGACCTGCTCTCTCTCTCACTGCTCAGGAACACAGGTGGAGTTCGGT 8685
Qy      2675 AspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAspProGlyThrVal 2694
Db      8686 GACCTGGGATCCCGGCTCATAGCATCCGTTTGGGGACACCTTGTATCAGGACCTGTG 8745
Qy      2695 MetArgPheSerCySerGlnAlaGlyHisValLeuArgIleSerSerGlyLysGln 2714
Db      8746 ATGCCCTTCACTGTGAAGTGGCCACAGTGTCCGGGATGTCAAGGCCGACCTGTCTCA 8805
Qy      2715 AlaAsnGlySerTrrSerGlySerGlnProGlnLysGlyValIleSerCySerGlyAsnPro 2734
Db      8806 GCCATGTCTGTGAGGCGCTCGACCTGAGTGTGAGATCTTGTGGGAACCT 8865
Qy      2735 GlyThrProSerAsnAlaArgValValPheSerAspGlyLeuValPheSerSerIle 2754
Db      8866 GGGACTCCAAGTATGCCCAGTGTGTTCAGTATGGCTGTCTCAGCTCTATC 8925
Qy      2755 ValTyrrGlnCyAsnGlnGlyTyrrTyrrAlaThrGlyLeuLeuSerArgHisCysSerVal 2774
Db      8926 GTCTATAGTGCCTGGGAAGATCTATACGACAGGCTGTCTGACGCTGCTCGCTC 8985
Qy      2775 AsnGlyThrTrrThrGlySerAspProGlnCysLeuValIleAsnCySerGlyAspProGly 2794
Db      8986 AATGTACTCGGACAGGCACTGACCTGATGCTCTGTCTATTAAGTGTGACCCCTGGG 9045
Qy      2795 IleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgTyrrAsnTyrrValThrTyrr 2814
Db      9046 ATTCCAGCCCAATGGCTTCCGCTGGGCAATGACTTCAAGGTAACAACAACTGTGACATAT 9105
Qy      2815 GlnCySerValProGlyTyrrMetMetGlnSerHisArgValSerValLeuSerCyThrLys 2834
Db      9106 CAGTGTCTCTGCTGTATATGATGAGTCAACATAGAGTATCTGTGTGAGCTGACCAAG 9165
Qy      2835 AspArgThrTrrPheGlnGlyThrLysProValCySerAlaLeuMetCysLysProProPro 2854
Db      9166 GACCGGACATGGAATGGAACCAAGCCGCTGGAAGCTCTCATGTGCAAGCACTCCG 9225
Qy      2855 LeuIleProAsnGlyLysValValGlySerAspPheMetTrrGlySerSerValThrTyrr 2874
Db      9226 CTATCCCAATGGAGGTGGTGTGATCTTCACTGTGGGGCTCAAGTGTGACTTAT 9285
```

QY 2875 AlaCsteuGluglYfYrGlnLeuSerLeuProAlaValPheThrCysGluGlyAspGly 2894
|||
Db 9286 GCCGCTCGAGGGGTTACCACTCTCCCTGCCCGGCTGTTCACTGTGAGGAAATGG 9345
|||
QY 2895 SerTrpThrGlyGluLeuProGlnCysPheProValPheCysGlyAspProGlyValPro 2914
|||
Db 9346 TCCTGACCCGAGAGCTGCTCACTGTTCTCTGTTCTCGGGGATCCGGTGTCCG 9405
|||
QY 2915 SerArgGlyArgGlyIuAspArgGlyPheSerTrpArgSerSerValSerPheSerCys 2934
|||
Db 9406 TCCCGTGGAGAGAGAGAGACCGAGCTTCTCTCAACAGTCATCTGCTCTCTCTCC 9465
|||
QY 2935 HisProProLeuValLeuValGlySerProArgArgPheCysGlnSerAspGlyThrTrp 2954
|||
Db 9466 CATCCCTCTGCTGCTGCTGCTGCTCTCCACGAGCTTTCCTGCTGAGTGGAGACATGG 9525
|||
QY 2955 SerGlyTrpGlnProSerCysAlaAspProThrLeuThrCysAlaAspProGlyVal 2974
|||
Db 9526 AGTGGACCCAGCCAGCTGATGATCCGACCTGACACGTTGGGACCTGCTGTG 9585
|||
QY 2975 ProGlnPheGlyIleGlnAsnAsnSerGlnGlyTrpGlnValGlySerThrValLeuPhe 2994
|||
Db 9586 CCAAGTTTGGGATACAGAACAATTCTCAGGGCTACAGGTTGAGACACAGTCTCTTC 9645
|||
QY 2995 ArgCysGlnIlysglyTrpLeuLeuGlnGlySerThrTrpArgThrCysLeuProAsnLeu 3014
|||
Db 9646 CGTTGTCAAAAGGCTACTGCTCTCAGGGCTCCACACCGAGACTGCTCTCCAAACCTG 9705
|||
QY 3015 ThrTrpSerGlyTrpProProAspCysValProHisIleCysArgGlnProGlyThrPro 3034
|||
Db 9706 ACCGAGATGGAGACCCACCACTGATGTGTCCCCACACATGCGACGACGAGACCCCA 9765
|||
QY 3035 ThrHisAlaAsnValGlyAlaLeuAspLeuProSerMetGlyTrpThrLeuIle-ThrPr 3054
|||
Db 9766 ACGATATCCACAGCTGGGGGCTCGATTTGGCTCCATGCGCTACAGCTCATCTACTCC 9825
|||
QY 3054 AlaArgArgAlaSerProSerArgValAlaProSerThrAlaProAlaArgArgMetAl 3074
|||
Db 9826 TGCAGAGAGGCTTCTCTCTCAAGGAGTGCTCCAGACACGACCTCCAGGCGGATGGC 9885
|||
QY 3074 aaIaGlyAlaSerArgProSerAlaTrpArgSerGlyProValGlyAspProSerTh 3094
|||
Db 9886 AGCTGACAGAGCAAGCCGCCATCTGCTGAGAGTCCGGCCATGGAGATCCATCAAC 9945
|||
QY 3094 rLeuProGlySerHisArgSerProIlyPro 3104
|||
Db 9946 ACTGCCCGGAGACCGCTACCCCAAGCCT 9976
|||
RESULT 7
ADH71143
ID ADH71143 standard; DNA; 10466 BP.
AC ADH71143;
DT 25-MAR-2004 (first entry)
DE Human gene of the invention NOV4 SEQ ID NO:39.
XX
XX de; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;
XX anorectic; antidiabetic; antimicrobial; antidiabetic; gene therapy;
XX KM vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX KM obesity; diabetes; infectious disease; metabolic syndrome X;
XX dyslipidaemia.
XX Homo sapiens.
XX OS
XX PN WC02003102155-A2.
XX
XX PD 11-DEC-2003.
XX
XX PF 03-JUN-2003; 2003WC-US017430.
XX

PR 03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 05-JUN-2002; 2002US-0386047P.
PR 06-JUN-2002; 2002US-0386376P.
PR 06-JUN-2002; 2002US-0386453P.
PR 06-JUN-2002; 2002US-0386864P.
PR 06-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386816P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0386942P.
PR 07-JUN-2002; 2002US-0386971P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-0386960P.
PR 10-JUN-2002; 2002US-0387400P.
PR 10-JUN-2002; 2002US-0387535P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388066P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390006P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 09-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 12-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403511P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 13-AUG-2002; 2002US-0406317P.
PR 15-AUG-2002; 2002US-0403617P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 17-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.

PR 05-NOV-2002; 2002US-00423798.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Albrock JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
 PI Cateron E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
 PI Ethenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
 PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li U, Liu X, MacDougall JR;
 PI MacLachlan T, Malyskar UM, Mezick AJ, Millet I, Mishra VS;
 PI Padigaru M, Raturajan M, Pena CE, Peyman JA, Raha D, Rastelli L;
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
 PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M,
 PI Zhong H;
 XX
 DR WPI: 2004-081935/08.
 XX P-PSDB; ADH71144.
 XX
 PT New NOXV polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOXV-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 4; SEQ ID NO 39; 1880bp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (NOXV). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antilipemic activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOXV polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing.
 CC CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC encodes a NOXV polypeptide of the invention.
 XX
 SQ Sequence 10466 BP; 2222 A; 3160 C; 2832 G; 2252 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 10466
 Score: 15917.50 Matches: 2946
 Percent Similarity: 95.37% Conservative: 21
 Best Local Similarity: 94.70% Mismatches: 45
 Query Match: 93.72% Indels: 100
 DB: 12 Gaps: 11
 US-10-016-248-2 (1-3104) x ADH71143 (1-10466)
 QY 2 AAGIYALAPROFROALALEULEUProCYSeSerLeuile---SerAspCys--- 19
 DB 731 TCTGGTTACCGGAGCGAGCCTCCGCCCTTATCGACGAGCAAACTGGTGGAC 790
 QY 20 CysAlSerAsnGlnIrhgIserValGlyValGly-----ProSerGluLeuVal 36
 DB 791 TGACTTTCACATGCGATGCGCAACCGCGACGCGGATTCAGTCCCAATGCCAA-ATC 849
 QY 37 LysLysGlnIleGluLeuLysSerArgValValLysLeuMetProSerLysAspAsnSer 56
 DB 850 AAGAGCAATGAGTTGAGTGAAGTCTCGAGGTGTGAAGCTGATGCCAGCAAAAGCAAGC 909
 QY 57 GlnLysThrSerValLeuThrGlnValGlyValSerGlnGlyLysAsnMetCysProAsp 76
 DB 910 CAGAGACCTCTGTGTACTCAGGTGGTGTGTCTCCAAAGACCAATATGTTCTCCAGC 969
 QY 77 ProGlyIleProGluAlaGlyLysArgLeuGlySerAspPheArgLeuGlySerVal 96
 DB 970 CTTGGCATATCCGAAAGGGGCAAAAGACTAGGCTCGATTTCAGGTGATCCAGCGTC 1029

QY 97 GlnPheThrCysAsnGluGlyLysAspLeuGlnGlySerLysArgLysThrCysMetLys 116
 DB 1030 CAGTTCACTGCAACGAGGGCTATGACTCGAAGAGGTCCAAAGCGATCACTCTATGAAA 1089
 QY 117 ValSerAspMetPheAlaAlaTrpSerAspHISArgProValCysArgAlaArgMetCys 136
 DB 1090 GTAGGAGCAATGTTTCCGCGCTCGAGCGACCAAGCCAGTCTGCGAGCCGCAATGTGT 1149
 QY 137 AspAlaHISLeuArgGlyLysProSerGlyIleIleThrSerProAspPheProIleGlnTyr 156
 DB 1150 GATGCCACCTTCGAGGCGCCCTCGGCAATCACTCCCAATTTCCCAATTCAGTAT 1209
 QY 157 AspAsnAlaHISCysValTrpIleIleThrAlaLeuAsnProSerLysValIleLys 176
 DB 1210 GACCAACATGACACACTGTGTGATCANTCAGCACTCAACCTCCCAAGTATATCAAG 1269
 QY 177 LeuAlaPheGluGluPheAspLeuGluArgGlyLysArgThrLeuThrValGlyAspGly 196
 DB 1270 CTGCGCTTGAAGAGTGTGATTTGAGAGGGGCTATGACACCTTCAGCGTCTGTATGCT 1329
 QY 197 GlyGlnAspGlyAspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSer 216
 DB 1330 GGTCAAGATGGGACCAAGACAGTCTCTATCATG----- 1365
 QY 217 ProHISThrProGlySerArgLysProGluSerMetSerGlyAspIleTrpArgLysLys 236
 DB 1366 -----CTGACAGGATCATCGGTCCCGATCTCATTTGCACACCAAT---CATCAATG 1416
 QY 237 TrpThrValLeuGluIleCysArgAspLysSerSerAspAlaArgSerGlySerVal 256
 DB 1417 TGGCTCTCTTCAG-----ACTGATGCC--AGTGGAGT--- 1449
 QY 257 ArgLysSerProLysThrSerAsnAlaValGluLeuValAlaProGlyThrGluIleGlu 276
 DB 1450 -----TCCCTGGATTCAGAGTCTCTTGAAGATATGAG 1485
 QY 277 GlnGlySerCysGlyAspProGlyIleProAlaTyrGlyLysArgGluGlySerArgPhe 296
 DB 1486 CAGGCACTTCGGTGAACCTTGCTGATACCTGATATGATGCGGAGGAGAGGCTCCGGT 1545
 QY 297 HISISGlyAspThrIleLysPheGluCysGlnProAlaPheGluLeuValGlyLys 316
 DB 1546 CGCCACGGTGACACATCAAGTTGATGTCAGCGCCGCTTATGCTGTGGACAAAG 1605
 QY 317 AlaIleThrCysGlnLysAsnAsnGlnTrpSerAlaLysLysProGlyCysValPheSer 336
 DB 1606 GCAATCAATGCGCAAAAGAAATTAACCAATGTCGCTAAGAAAGCAGGCTCGTGTTC 1665
 QY 337 CysPhePheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAsp 356
 DB 1666 TGCTTCTTCAACTTCAACCAAGCCGCTCGGGTGTCTGTCTTCCCACTAACCAAGAGAC 1725
 QY 357 TyrGlyAsnHISLeuHISCysValTrpLeuIleLeuAlaArgProGluSerArgLysHIS 376
 DB 1726 TATGGCAACACCTTCCACTGTGTCTGCTCATCTGCGCAGGCTGAGGCGGCATATCC 1785
 QY 377 LeuAlaPheAsnAspIleAspValGluProGlnPheAspPheLeuValIleLysArgGly 396
 DB 1786 CTGGCTTCAACACATTTGAGTGAAGCTCAAGTATGATTTCTGTGTCAATCAAGATGGG 1845
 QY 397 AlaThrAlaGluAlaProValLeuGlyLysThrPheSerGlyAsnGlnLeuProSerSerIle 416
 DB 1846 GCCACCGCGAGGCGCGCGTCTGGGCACTTCTCAAGAAACAGTTCCTCTCTCATC 1905
 QY 417 ThrSerSerGlyHISValAlaArgLeuGluPheGlnThrAspHISerThrGlyLysArg 436
 DB 1906 ACAAGCAGTGGCGACCTGCGCTCGAGTTCAGATTCAGACCACTCCACAGGAAAGAG 1965
 QY 437 GlyPheAsnIleThrPheThrThrPheArgHISAsnGluCysProAspProGlyValPro 456
 DB 1966 GCGTTCACATCACTTTTCAAGTATGATCTCCAAAGATGCGCGGATTCCTGCGTCCA 2025
 QY 457 ValAsnGlyLysArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCys 476

```
Db 2026 GTAAATGGCAAAAGCGTTTGGGGAGACGCTCCAGCTGGGCACTCATCTCTCTCTGT 2085
Qy 477 AepgluglyPheLeuGlyThrGlnGlySerGluThrIleThrCysValLeuGlyGluY 496
Db 2086 GATAAAGGCTTCCTTGGGACTCAGGGCTCAGAGACATCACCTGCTCTGAAAGAGGCG 2145
Qy 497 SerValVal1TrpAnserAlaValLeuArgCysGluAlaProCysGlyGlyYhiLeuThr 516
Db 2146 AGCGTGTCTGAAACAGCGCTGTGTGCGGTGAAGTCCCTGTGGGTGCACCTGACT 2205
Qy 517 SerProserGlyThrIleLeuSerProGlyTrpProGlyPheYrYLyAspAlaLeuSer 536
Db 2206 TCGCCAGCGGACCATCTCTCTCCGGGCTGGCTGGCTCTTCAAGAGAGCTTGAAC 2265
Qy 537 CysAlaTrpVal11IleGluAlaGlnProGlyYrYrProIleYsIleThrPheAspArgPhe 556
Db 2266 TGTGCTGTGGTGAATGAGGCCAGCCAGGCTAACCCCATCAAAATCACCTTCGACAGGTTG 2325
Qy 557 LyeThrGluValaAnrYrAspThrLeuGluValaArgAspGlyArgThrYrSerAlaPro 576
Db 2326 AAAACCGAGGTCACTATGACACCTGGAGTACGCGATGGCGGACTTACTCAGCGCC 2385
Qy 577 LeuIleGlyVal1YrYhiGlyThrGlnVal1ProGlnPheLeuIleSerThrSerAsnYr 596
Db 2386 TTGATCCGGGCTTACACAGGAGCCAGGTTCCCAATTCCCTCATCAGACAGCACTAC 2445
Qy 597 LeuYrLeuLeuPheSerThrAspYrYsSerHisSerAspIleGlyPheGlnLeuArgYr 616
Db 2446 CTCTACCTCCCTCTCTTACCGACAGACAGTCACTCGGACATCGGCTTCCAGCTCCGCTAT 2505
Qy 617 GluThrIleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGln 636
Db 2506 GAGGCTTAAACACGTGACGTACAGACCATGTGATCCAGGAATCCAGTAAAGGACAG 2565
Qy 637 ArgHisGlyAsnAspPheYrVal1GlyAlaLeuVal1ThrPheSerCysAspSerGlyYrY 656
Db 2566 CGTATGGGATGACTTCTTACGTGGGCGCGCTGGTGAACCTTCACTGTGACTCGGGCTAC 2625
Qy 657 ThrLeuSerAspGlyGluProLeuGluCysGluProAsnPheGlnTrpSerArgAlaLeu 676
Db 2626 ACATTAAAGTACGGGGAGGCTCTGGAGTGTAGCCCAACTTCCAGTGGAGCGGGCCCTG 2685
Qy 677 ProserCysGluAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyYhiLeuSer 696
Db 2686 CCCAGTTGTGAAGCTCTGTGGTGGCTTCAATCAAGGCTCCAGTGGAGCCATCTTTCG 2745
Qy 697 ProGlyPheProAspPheYrProAsnAsnLeuAsnCysThrTrpIleIleGluThrSer 716
Db 2746 CCAAGTTTCCCTGACTTCTTACCCCAACACTTGAATGCACTGGATTATCGAAACATCT 2805
Qy 717 HisGlyYrGlyVal1PhePheThrPheHisThrPheHisLeuGlnSerGlyYhiAspYr 736
Db 2806 CATGGCAAGGGGTGTCTTCTTCACTTTCACACTTTCACCTGGAAGGGCCATGATGATC 2865
Qy 737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db 2866 CTCTCTCATCATGGAACGAGGAGCTTCAACCGAGCCCTGAGGAGCTTAACTGGATCTCGG 2925
Qy 757 LeuProAlaProIleSerAlaGlyLeuYrGlyAsnPheThrAlaGlnValArgPheIle 776
Db 2926 CTGCGACAGTCCCATCAGCGCTGGGCTCTATGGCAACTTCACTGCCCGAGTCCGTTATC 2985
Qy 777 SerAspPheSerMetSerYrGluGlyPheAsnIleThrPheSerGlyYrAspLeuGlu 796
Db 2986 TCTGATTTCTTCATGTCTATGAGATTAACAATCACTTCTCAGAGTACGACTTGGAG 3045
Qy 797 ProCysGluGluProGluVal1ProAlaYrSerIleArgGlyGlyLeuGlnPheGlyVal 816
Db 3046 CCCGTGAGAGGCCGAGGCTCCAGCCTACAGCATCCGGAAGGGGCTTGCAGTTTGGGCTG 3105
Qy 817 GlysAspThrLeuThrPheSerCysPheProGlyYrYrArgLeuGluGlyYhiAlaArgIle 836
Db 836 GGCAGACACTTGAACCTTCTCTGCTTCCCGGGTACCCTGAGGAGCCGCGCATC 3165
Qy 837 ThrCysLeuGlyGlyAlaArgArgLeuTrpSerSerProLeuProArgCysValAlaGlu 856
Db 3166 ACGTGCCTGGGGGAGACCGCGCTGTGAGGCTCGCTTGCACAGGTGTGTTGCTGAG 3225
Qy 857 CysGlyAsnSerVal1ThrGlyThrGlnGlyYhiLeuLeuSerProAsnPheProValAsn 876
Db 3226 TGTGGGATTCAGTCAAGGACTCAAGGATCACTTGTGTGTCGCCCACTTTCCTGTAAAC 3285
Qy 877 TyrAsnAsnAsnHisGlyCysIleYrYSerIleGlnThrGlnProGlyYrGlyIleGln 896
Db 3286 TACAATTAACAATATGATGATCATCTACTTCACTCAGACCCAGCAGGAGGAATTCAG 3345
Qy 897 LeuYrAlaArgAlaPheGluLeuSerGlnGlyYrAspVal1LeuYrVal1YrAspGlyAsn 916
Db 3346 CTGAAGCCAGGGGATTCGAATCTCCGAAGAGAGATCTCTCAAGGTTTATGATGGCAGC 3405
Qy 917 AsnAsnSerAlaArgLeuLeuGluVal1PheSerHisSerGlyLeuMetGlyVal1ThrLeu 936
Db 3406 AACACTCCCGCGTTGTGTGGAGTGTATTAAGCCATTCAGATGATGGGGGTGACTTGTG 3465
Qy 937 AsnSerThrSerSerSerLeuTrpLeuAspPheIleThrAspAlaGluAsnThrSerYs 956
Db 3466 AACAGCATCCAGGACGTCTGTGGCTTGAATTCATCATCTATGCTGAATAAACACAGCAG 3525
Qy 957 GlyPheGluLeuHisPheSerSerPheGluLeuIleYsCysGluAspProGlyYhiPro 976
Db 3526 GGCCTTGAACGTGACCTTTTCCAGTTTGAATCATCAATCAATGATGAGGACCCAGAACCC 3585
Qy 977 LysPheGlyYrYrYsVal1HisAspGluGlyYhiPheAlaGlySerSerVal1SerPheSer 996
Db 3586 AAGTTTGGCTTAAAGGTTCAAGATGAAGAGTCAATTTGACAGGAGCTCCGTGTCTTACG 3645
Qy 997 CysAspProGlyYrYrSerLeuAspArgYrYrSerGluGluLeuLeuYsLeuSerGlyGluArg 1016
Db 3646 TGTGACCTCGATACAGCCCTGCGGGGTATGAGAGAGCTGTGTGTGATGAGAGAGCCG 3705
Qy 1017 ArgThrTrpAspArgProLeuProThrCysValAlaGluCysGlyGlyYhiArgGly 1036
Db 3706 CGGACCTGGGAGCCGCGCTTCCACCTGTGTGCGCCAGGTGAGGAGGACAGTGAAGAGA 3765
Qy 1037 GluValSerGlyGlnVal1LeuSerProGlyYrYrProAlaProYrGluHisAsnLeuAsn 1056
Db 3766 GAGGTGTGGGGAGGAGGTGTGTACCCGGGTATCCAGCTCCATGAAACAATCTTAAC 3825
Qy 1057 CysIleTrpThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPhe 1076
Db 3826 TGCATCTGACCATCCAGAGAGAGCCGCTGCAACATTCGGCTACACTTCCGTGTCTT 3885
Qy 1077 AspThrGluGluVal1HisAspVal1LeuArgIleTrpAspGlyProValGluSerGlyVal 1096
Db 3886 GACACAGAGGAGGTTTCAAGAGTGTGCGCATCTGGGATGGGCTGTGGAGAGAGGGGTT 3945
Qy 1097 LeuLeuLeuGluLeuSerGlyProAlaLeuProYrAspLeuHisSerThrPheAsnSer 1116
Db 3946 CTGCTGAAGAGGTGATGGTCCGCGCCCTGAGGACCTGTGATGACACTTCACTCG 4005
Qy 1117 ValValLeuGlnPheSerThrAspPhePheThrSerYrGlnGlyPheAlaIleGlnPhe 1136
Db 4006 GTGCTCTGAGATTCAGACATGACTTCTTCAACAGAGAGAGGCTTTCGCAATTCATTT 4065
Qy 1137 SerValSerThrAlaThrSerCysAsnAspProGlyYrLeuProGlnAsnGlySerArgSer 1156
Db 4066 TCAGGTTCCACAGCAACGTCTCTCAATGACCTCGGATCCCGAGAAATGGAGTCCGAGT 4125
Qy 1157 GlysAspSerTrpGluAlaGlyAspSerThrVal1PheGlnCysAspProGlyYrYrAlaLeu 1176
Db 4126 GTGTACAGTTGGGAAGCGCGGAGCTCCACAGTGTTCAGTGTACCTGTGCTAGCGGCTG 4185
Qy 1177 GlnGlySerAlaGluIleSerCysVal1YsIleGluAsnArgPhePheTrpGlnProSer 1196
Db 4186 CAGGGAAGTCCAGAGATCAGCTGTGTGAAGATGAGAAACAGGTTCTTTCGCGAGCCGAGC 4245
```

QY	1555	CYSGIYGLYASNIleuThrgIuArygrArgIYThrIIeuserProGlyIphProGluPro	1574
Db	5526	TGTCGAGCAACTCTCAAGAGCGAGGGCACATCTGTCCCTCGGCTTCCCAAGGCCG	5385
QY	1575	TYTleuASerIeuanSCysValITrpySLIeValIProGluGlyAIgIYIIeGln	1594
Db	5386	TACCTCAACGCTCACTGTGTGTGGAAAGATCGTGTCCCGAAGCGCTGGCATCCAG	5445
QY	1595	IIAGIIValValSerPheIalThrgIuGlnIleuThrPaSerSerLeuGluValPheApGly	1614
Db	5446	ATCCAAgTGTCAgTGTGTGTGCAAGAGCAAGACTGGGACTGCTGGAAgTATTGATGTGT	5505
QY	1615	AlaASpASnThrValThrMetLeuGlySerPheSerGlyThrThrValProAlaIleuLeu	1634
Db	5506	GCAGATPAACGTGTAAACCATGCTGGGAGCTTCTCAGGAACAACTGCTGCTCCCTTGTG	5565
QY	1635	ASnSerThrSerASnGlnLeuTYrIleuHISpHeTYrSerApIIeSerValSerAlaIa	1654
Db	5566	AAACAGACCTCCAAACAGCTCACTTCATTTCACTACAGATATCAAGGTATCTGCAGCT	5625
QY	1655	GIYPheHISLeuGluTYrLeuThrValGlyIleuserSerCysProGluProAlaValPro	1674
Db	5626	GGCTTCCACTTGGAGTACAAAACGTGGGCTTGAGCAgTGTCTCGAAACCTGCTGTGCC	5685
QY	1675	SerASnGlyValIleThrGlyIuArgTYrIleuValASnApValIaISerPheGlnCys	1694
Db	5686	AGTAAACGGGTGTAAAGACTGCGAGCGCTACTTGTGTATGATGTGTGTCTTCCAGTGT	5745
QY	1695	GIUPProGlyTYrAlaIleuGlnGlyHISAlaHISIIeSerCysMetProGlyThrValArg	1714
Db	5746	GAGCCGGAGATAGCCCTTCAGAGGCGACGCCCATCTCCGCAATGCGCCGGAACAGTGGCGG	5805
QY	1715	ArgTYrASnTYrProProProLeuCysIleAlaGlnCysGlyTYrThValGluGluMet	1734
Db	5806	CGATGAAACTACCTCCCTCCACTGTGATTTGCAACAGTGTGGGGAACAGTGGAGGAGATG	5865
QY	1735	GluGlyValIIleIleuserProGlyIphProGlyASnTYrProSerASnMetAcPcySer	1754
Db	5866	GAGGGGGGATCTCTGAGGCCGCGGCTTCCCAAGGCAACTACCCAGTAACTGAGCTCTCCG	5925
QY	1755	TrpySLIeAlaIeProValGlyIphGlyAlaHISIIeGlnPheIleuASnPheSerThr	1774
Db	5926	TGGAAAATAGCACTGCCCGTGGCTTGGAGCTCAATCCAGTCTCTGAACTTCTCCACC	5985
QY	1775	GluProASnHISApTYrIIeGluIIeArgASnGlyProTYrGlnThrSerArgMetCet	1794
Db	5986	GAGCCCAACCCGACTCATAGAATCCGGAAATGGCCCTCATAGAACCGCGCATGATG	6045
QY	1795	GlyArgPheSerGlyIserGluASnProSerSerIleIleuserThrSerHISGluThrThr	1814
Db	6046	GGAAGATTCAGTGGAAACGAGCTTCCAAAGTCTCTCTCTTCCACGTGCCACAGAACACC	6105
QY	1815	ValTYrPheHISerASpHISerGlnASnArgProGlyIphIleuLeuGluTYrGlnAla	1834
Db	6106	GTCATATTCCACAGCGACCACTCCAGAAATCGCGACAGATTCACAGCTGAGATTCAGGCC	6165
QY	1835	TYrGluIleuGlnGlyCysProApProGluProPheAlaASnGlyIleValIArgGlyAla	1854
Db	6166	TATAACTTCAAGAGTCCCAAGCCCAAGAGCCCTTTGGCCAAATGGCATGTGTGAGGGAGCT	6225
QY	1855	GlyTYrASnValGlyIleuserValThrPheGluCysIleuProGlyTYrGlnIleuThrgIy	1874
Db	6226	GGCTACAACTGGGACATACAGACTTTCGAGTGCCTCCCGAGGTATCAATTGACTGC	6285
QY	1875	HISProValIleuThrCysGlnHISgIyThrASnArgASnTYrPASPHisProIleuProIys	1894
Db	6286	CACCTGTCTCTCACTGTCAACTGAGCAACCAACCGAACTGGAGCAACCCCTGCGCAAG	6345
QY	1895	CysGluValAProCysGlyIyIyASnIleThrSerSerASnGlyThrValTYrSerProGly	1914
Db	6346	TGTGAAGTCCCTTGTGTGGGAAACATCACTTCTTCCAAACGGAACGTGTACTCCCGGGG	6405
QY	1915	PhaProSerProTYrSerSerSerGlnIleASnCysValITrPLeuIleThrValProIIeGly	1934

```
Db 6406 TTCCCTACCCGCTACTCCAGCTCCAGGACGTGTCTGCTGATCAGCGTCCATTGGC 6465
Qy 1935 HtbcglValaIargLeuAanLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThr 1954
Db 6466 CATGGCCGTCGCCCTCAACCTCAGCTGTGACAGACAGCCCTCGGAGATTTTCATCC 6525
Qy 1955 IletPaspArgLyProGlnGlnThrAlaProArgLeuGlyValaPheThrArgSerMetAla 1974
Db 6526 ATCTGGATGGGCCACAGCAACACACACGCTCGCGCTTTCACCCGGAGCATGGCC 6585
Qy 1975 LysylThrValaGlnSerSerSerAangGlnValLeuLeuLysPheHisArgAspAlaAla 1994
Db 6586 AAGAAAACAGTGCAGAGATTCAATCCACAGGTCCTGCTCAAGTTCCACCGATGACAC 6645
Qy 1995 ThrGlyGlyIlePheAlaIleAlaPheSerAlaTyProLeuThrLysCysProProPro 2014
Db 6646 ACAGGGGGGATCTTCGSCATAGCTTTCCTCCCTATCCACTCACCAATGCCCCCTCC 6705
Qy 2015 ThrIleLeuProAanAlaGluValaValaThrGluAangGlnLupPheAanIleGlyAspIle 2034
Db 6706 ACCATCTCTCCCAACGCGGAAGTGTCAAGAGATGAAGATTCATATAGTGCATC 6765
Qy 2035 ValArgTyArgCysLeuProGlyPheThrLeuValaGlyAangGlnIleLeuThrCysLys 2054
Db 6766 GTAGGCTACAGATGCTCCCTGGCTTTACTTAACTAGGGGAATGAATTTCTACCTGCAA 6825
Qy 2055 LeuGlyThrTyLeuGlnPheGlnGlyLupProProIleCysGluValaHisCysProThr 2074
Db 6826 CTGGAACTTACCTGACGTGTGAAGGACACCCCGCATATGTGAAGTCACTGTCACAA 6885
Qy 2075 AangIleLeuLeuThrAaspSerThrGlyValaIleLeuSerGlnSerTyProGlySerTy 2094
Db 6886 AATAGCTTCTGACAGACTCCACAGGCGTATCTTGACCCAGAGTACCTCGAAGCAT 6945
Qy 2095 ProGlnPheGlnThrCysSerTrpLeuValaArgValaGluProAspTyAsnIleSerLeu 2114
Db 6946 CCCAGTTCACAGACTGCTCTGGCTGTGAGAGTGCAGCCGACTAATACCTCCCTC 7005
Qy 2115 ThrValaGlyTyPheLeuSerGlnLysGlnTyArgGluPheGluIlePheAspGlyPro 2134
Db 7006 ACAGTGGAGTACTCTCAGCGAGAACATATGATGTGATTTTGATGTGATGCTCA 7065
Qy 2135 SerGlyGlnSerProLeuLeuValaLeuSerGlyAanTyArgSerAlaProLeuIleVal 2154
Db 7066 TCAGGACAGAGTCTCTGCTAAGGCTCAGTGGAAATTACTACGCTCCCTATGTC 7125
Qy 2155 ThiserSerSerAanSerValTyLeuArgTrpSerSerAaspHisAlaTyArgAspGly 2174
Db 7126 ACCAGCTCAACCACTGTGTACTGCTGCTGCTGCTATCTGATCAAGCTTCAATCGAAG 7185
Qy 2175 GlyPheLysIleArgTySerAlaProTyCysSerLeuProArgAlaProLeuHisGly 2194
Db 7186 GGGCTTCAAGATCCGCTATTCA----- 7206
Qy 2195 PheIleLeuGlyGlnThrSerThrGlnProGlyGlySerIleHisPheGlyCysAanAla 2214
Db 7206 ----- 7206
Qy 2215 GlyTyArgLeuValaGlnHisSerMetAlaIleCysThrArgHisProGlnGlyTyHis 2234
Db 7206 ----- 7206
Qy 2235 LeuTrpSerGlnAlaIleProLeuGlyGlnAlaLeuSerCysGlyLeuProGlnAlaPro 2254
Db 7207 -----GCCCTTCTCTGTGGGCTTCTCGAGGCCCCC 7236
Qy 2255 LysAangLysMetValPheGlyLysGlyLupTyThrValaGlyThrLysAlaValaTySerCys 2274
Db 7237 AAGAAATGGAATGGCTTTGGCAAGGAGTACACAGTGGGACCAAGGCCATGTACAGCTGC 7296
Qy 2275 SerGlyGlyTyHisLeuGlnAlaGlyValaGlnAlaThrAlaGluCysLeuAaspThrGly 2294
Db 7296 ----- 7296
Db 7297 AGTAAAGGCTTACCACTCCAGGCGAGGCTGAGGCCACTGACAGATGCTTGACACAGGC 7356
Qy 2295 LeuTrpSerAanAanValaProProGlnCysValaProValaThrCysProAaspValSer 2314
Db 7357 CTATGGAGCAACCGCAATGTCCACACAGTGTCTCTGTGACTTGTCCGAAATGCAT 7416
Qy 2315 SerIleSerValaGlnHisGlyArgTrpArgLeuIlePheGluThrGlnTyGlnPheGln 2334
Db 7417 AGCATACGCTGGAGCATGGCCGATGGAGGCTTATCTTTAGACACAGTACATTCAG 7476
Qy 2335 AlaGlnLeuMetLeuIleCysAaspProGlyTyTyTyTyThrGlyGlnArgValaIleArg 2354
Db 7477 GCCAGCTGATGCTCATCTGTGACCTGGCTACTACTACTGSCAAAGGCTCATCGC 7536
Qy 2355 CysGlnAlaAangGlyLupTrpSerLeuGlyAaspSerThrProThrCysArgIleIleSer 2374
Db 7537 TGTACAGGCAATGGCAATGGAGGCTGGGAGCTTACGCCACCTCCGCAATCATCTCC 7596
Qy 2375 CysGlyGluLeuProIleProProAanGlyHisArgIleGlyThrLeuSerValTyGly 2394
Db 7597 TGTGAGAGCTCCCGATTTCCCAATGGCCACCGCATCGGAACACTGTCTGTACGG 7656
Qy 2395 AlaThrAlaIlePheSerCysAanSerGlyTyThrLeuValaGlySerArgValaArgGlu 2414
Db 7657 GCAACAGCATCTTCTCTGCAATTCGAGATACACATGCTGAGGCTCAGGGGTGCTGAG 7716
Qy 2415 CysMetAlaAangGlyLeuTrpSerGlySerGlyValaArgCysLeuAlaGlyHisCysGly 2434
Db 7717 TGCATGGCCAAATGGGCTCTGGAGTGTGAAATCGCGCTGCTGTGACACTGTGG 7776
Qy 2435 ThrProGluProIleValaAangGlyHisAlaAangGlyLupAanTyArgGlySer 2454
Db 7777 ACTCTGAGCCCATTTGCAACGACACACATCATATGGAGAACTACAGTACCGGGAGCT 7836
Qy 2455 ValValaTyGlnCysAanAlaGlyPheArgLeuIleGlyMetSerValaArgIleCysGln 2474
Db 7837 GTGGTGAACCATGCAATGCTGGCTTCGCTCGCTGAGCATGTGTGGCATGTGCAG 7896
Qy 2475 GlnAaspHisTrpSerGlyLysThrProPheCysValaProIleThrCysGlyHisPro 2494
Db 7897 CAGGATCATCACTGTGGGCGAAGACCCCTTCTGTGTGCCAATTAACCTGTGGACACCA 7956
Qy 2495 GlyAanProValaAangGlyLeuThrGlnGlyAangGlnPheAanLeuAanAanValaLys 2514
Db 7957 GGCACACCTGTCAACGCTCTCATCAGGGTACCAAGTTTAACCTCAACGATGTGTACAG 8016
Qy 2515 PheValaCysAanProGlyTyMetAlaGluGlyAlaAlaArgSerGlnCysLeuAlaSer 2534
Db 8017 TTTGTTTGCAACCTGGGTATATGCTGAGGGGGCTGTAGTCCCAATGCTGTGGCAGC 8076
Qy 2535 GlyGlnTrpSerAaspMetLeuProThrCysArgIleIleAanCysThrAapProGlyHis 2554
Db 8077 GGGCAATGAGTGCATAGCTGCCACCTGCAAGATCACTCAACTGTACAGATCCCTGGACAC 8136
Qy 2555 GlnGluAanSerValaArgGlnValaHisAlaSerGlyProHisArgPheSerPheGlyThr 2574
Db 8137 CAAGAAATATGTTGTCATAGTTCACGCAAGCGGCCGCAAGGTTTCAAGTTCGGACCC 8196
Qy 2575 ThrValSerTyArgCysAanHisGlyPheTyThrLeuGlnGlyThrProValaLeuSerCys 2594
Db 8197 ACTGTGTCTTACCGGTCAACCAAGGCTTACTCTGTGGGACCCCAAGTGTCAAGTGC 8256
Qy 2595 GlnGlyAaspGlyThrTrpAaspArgProArgProGlnCysLeuLeuValaSerCysGlyHis 2614
Db 8257 CAGGAGATGAGCATGAGGACCGTCCCGGCCCAAGTGTCTCGGTGTCTGTGGCAT 8316
Qy 2615 ProGlySerProProHisSerGlnMetSerGlyAaspSerTyThrValaGlyAlaValaVala 2634
Db 8317 CCGGGCTCCCGCTCACTCCAGATGTCTGGAGACAGTTATACTTGGAGACAGTGTGT 8376
Qy 2635 ArgTySerCysAlaGlyLysArgTyThrLeuValaGlyAanSerThrArgMetCysGlyLeu 2654
Db 8377 CGGTACAGCTGCAATCGCAAGCGTACTGTGTGGAAACAGCACCCGCAATGTGTGGCTG 8436
```

QY	2655	AspGlyIhSTPrThrGlySerLeuProHisCysSerGlyThrSerValGlyValCysGly	2674
Db	8437	GATGACACATGGACCTGGCTCCCTCCCTCACTGCTCAGAAACCAAGCGTGGAGATTGGCGT	8496
QY	2675	AspProGlyIleProAlaHisGlyIleArgLeuGlyAspSerPheAspProGlyThrVal	2694
Db	8497	GACCTGGGATCCGGCTCATGGCATCCGTTGGGGAGACAGTTGATCCAGGCACTGTG	8556
QY	2695	MetArgPheSerCysGluAlaGlyHisValLeuArgGlySerSerGlyIleArgThrCysGln	2714
Db	8557	ATGGCTTCACGCTGTGAAGCTGGCCACGATGCTCCGGGGATCTCTAGACGCACCTGTCA	8616
QY	2715	AlaAsnGlySerTrpSerGlySerGlnProGluCysGlyValIleSerCysGlyAsnPro	2734
Db	8617	GCCATATGCTCGTGGAGCGGCTCGACGCTGATGTGGAGATCTCTTGGGAACCT	8676
QY	2735	GlyThrProSerAsnAlaArgValValPheSerAspGlyLeuValPheSerSerIle	2754
Db	8677	GGGACTCCAGTAAATCCCGAAGTTGGTTCACTGATAGGCTGGTTTCTCCAGCTCATC	8736
QY	2755	ValTyrGluCysArgGluGlyTyrTyrAlaIhrgIleuLeuSerArgHisCysSerVal	2774
Db	8737	GCTATAGAGCCGGGAGAGATACTACGCCACAGGCTGCTCAGCCGCACTGCTCGCTC	8796
QY	2775	AsnGlyThrTrpThrGlySerAspProGluCysLeuValIleAsnCysGlyAspProGly	2794
Db	8797	AATGATACCTCGACAGACGACGTAACTCGATGCTCTGTCATAAACTGTGGTAACTCCGCG	8856
QY	2795	IleProAlaAsnGlyLeuArgLeuGlyAsnAspPheArgTyrAsnIleTrpValIhTrpTyr	2814
Db	8857	ATTCGACCAATGCTGCTGCGCTGGGCAATGACTTCAGGTACAAACAACTGGACATAT	8916
QY	2815	GlnCysValProGlyTyrMetMetClnSerHisArgValSerValLeuSerCysTrhTlys	2834
Db	8917	CAGTGTGCTCCGGCTCATATGATGGAGTCAATAGATCTGTGGCTGAGACTCACCAAG	8976
QY	2835	AspArgThrTrpAsnGlyThrIleAspProValCysIleValLeuMetCysIleAspProPro	2854
Db	8977	GACCGGACATGGAATGGAAACCAAGCCCTCTGCAGAAAGCTCTCATGTGCAGAGCCACTCG	9036
QY	2855	LeuIleProAsnGlyIleValValGlySerAspPheMetTrpGlySerSerValIhTrpTyr	2874
Db	9037	CTCATCTCCCAAGGGAGAGATGATGGGGTCTGACTTCAATGGGGCTCAAGTGTCACTTAT	9096
QY	2875	AlaCysLeuGluGlyTyrGlnLeuSerLeuProAlaValPheTrpCysGluGlyAsnGly	2894
Db	9097	GGCTGCTCGAGAGGGTATCAACTCTCCCTGCGCGGGTGTTCACCTGTAGGGAAATGGG	9156
QY	2895	SerTrpThrGlyGluLeuProGlnCysPheProValPheCysGlyAspProGlyValPro	2914
Db	9157	TCTTGACACGGAGAGCTGCTCATGTTCCTCGTGTTCGCGGGATCTCTGTGTCCG	9216
QY	2915	SerArgGlyArgArgGlyIleAspArgGlyPheSerTyrArgSerSerValSerPheSerCys	2934
Db	9217	TCCGTGGAGAGAGAGAGACCGAGGCTTCTCTTAAGGCATCTGTCTCTTCTCTGC	9276
QY	2935	HisProPheLeuValLeuValGlySerProArgPheCysGlnSerAspGlyIhTrp	2954
Db	9277	CATCCCTCTGTGTGGTGGGCTCTCCACGAGTTTTCGACATGCAATGGACATAGG	9336
QY	2955	SerGlyIhrgIhProSerCysIleAspProThrLeuThrThrCysAlaAspProGlyVal	2974
Db	9337	AGTGGACCCGACCCACTGCATATGATCCGACCTGACACAGTGTGGGACCTGTGTG	9396
QY	2975	ProGlnPheGlyIleGlnAsnAsnSerGlnGlyTyrGlnValGlySerThrValLeuPhe	2994
Db	9397	CCACAGTTTGGGATACAGACAAATTTCTCAAGGCTTACAGTTGGAAAGACAGTCTCTTC	9456
QY	2995	ArgCysGlnIleGlyTyrLeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeu	3014
Db	9457	CGTGTCAAAAAGGCTACTGTTCTCAGGGGTCCACACACAGAACTGCTCTCCAAACCTG	9516

QY	3015	ThrlrpserclyThrProProlaPpCyvsVAlProhshIsCybaArgGlnProGlnuThrPro	3034
Db	9517	ACCGGAGTGAACCCACCTGACTGTGTCTCCCCACCACTGCAGGACCCAGAGACGCCA	9576
QY	3035	ThrlhslAAsnValGlYAlaLeuAAspLeuProSerMetGlYTrpThleuIle-ThrPr	3054
Db	9577	ACCGATCCAACTCGGGGCCCTTGATTTGGCTTCATGGGCTACACGCTCATTTACTCC	9636
QY	3054	OhlaArgArgAlaSerProSerArgValAlaProSerThrlAProAlaArgMetaI	3074
Db	9637	TGCAGAGGGCTTCTCTCCCAAGGTGGCTCCGAGCACCCGACACTCGCAAGCGGATGC	9696
QY	3074	aAlaGlyGlnAlaSerArgProSerAlaTrpArgSerGlyProValAlaGlyAspProSerTh	3094
Db	9697	AGCTGACAGCAGCAGACCGCCCATCTGCTCGAGGTCTCGGCCCAATGGAGACCATCAAC	9756
QY	3094	rIeupProGlySerThrlsArgSerProLysPro	3104
Db	9757	ACTGCCGGAGGCCACCGCTCACCAACCT	9787
RESULT 8			
ID	ADH71145	standard; DNA; 12900 BP.	
XX	ADH71145;		
XX	25-MAR-2004	(first entry)	
XX			
DE	Human gene of the invention NOV4f SEQ ID NO:41.		
XX			
KM	de; gene; human; cytosolic; immunomodulator; neuroprotective; nootropic;		
KM	anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;		
KM	vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;		
KM	obesity; diabetes; infectious disease; metabolic syndrome X;		
KM	dyslipidaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2003102155-A2.		
XX			
PD	11-DEC-2003.		
XX			
PF	03-JUN-2003; 2003WO-US017430.		
XX			
PR	03-JUN-2002; 2002US-0385120P.		
PR	04-JUN-2002; 2002US-0385784P.		
PR	05-JUN-2002; 2002US-0386041P.		
PR	06-JUN-2002; 2002US-0386047P.		
PR	06-JUN-2002; 2002US-0386376P.		
PR	06-JUN-2002; 2002US-0386453P.		
PR	06-JUN-2002; 2002US-0386864P.		
PR	06-JUN-2002; 2002US-0387016P.		
PR	07-JUN-2002; 2002US-0386796P.		
PR	07-JUN-2002; 2002US-0386816P.		
PR	07-JUN-2002; 2002US-0386931P.		
PR	07-JUN-2002; 2002US-0386942P.		
PR	07-JUN-2002; 2002US-0386971P.		
PR	08-JUN-2002; 2002US-0387262P.		
PR	08-JUN-2002; 2002US-0296960P.		
PR	10-JUN-2002; 2002US-0387400P.		
PR	10-JUN-2002; 2002US-0387535P.		
PR	11-JUN-2002; 2002US-0387610P.		
PR	11-JUN-2002; 2002US-0387625P.		
PR	11-JUN-2002; 2002US-0387634P.		
PR	11-JUN-2002; 2002US-0387668P.		
PR	11-JUN-2002; 2002US-0387696P.		
PR	11-JUN-2002; 2002US-0387702P.		
PR	11-JUN-2002; 2002US-0387836P.		
PR	11-JUN-2002; 2002US-0387859P.		
PR	12-JUN-2002; 2002US-0387933P.		
PR	12-JUN-2002; 2002US-0387934P.		
PR	12-JUN-2002; 2002US-0387960P.		
PR	12-JUN-2002; 2002US-0388022P.		

PR 12-JUN-2002; 2002US-0386096P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389128P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389723P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390066P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 09-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 15-AUG-2002; 2002US-0403617P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 23-SEP-2002; 2002US-0412528P.
PR 30-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Allobrook JP, Alvarez E, Anderson DM, Boldog FI, Caeman SJ;
PI Carleton E, Chapoval A, Crabtree-Bokor JR, Edinger SK, Ellerman K;
PI Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
PI Guev VY, Hartmann JL, Ji W, Kekuda R, Li L, Liu X, Macdonald JR;
PI MacLachlan T, Malynkar UM, Mezick AJ, Millet I, Mishra VS;
PI Padigaru M, Paturajan M, Pena CE, Peyman JA, Raha D, Rastelli L;
PI Rigger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
PI Smithson G, Spytek KA, Stone DU, Vernet CM, Voss EZ, Zhong M,
PI Zhong H;
XX
DR WPI: 2004-081935/08.
DR P-PSDB; ADH71146.
XX
PT New NOXV polypeptides and nucleic acid molecules useful for preventing or
PT treating NOXV-associated disorders, e.g. cancer, diabetes, infection or
PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
XX
XX Example 4; SEQ ID NO 41; 1880bp; English.

CC The invention relates to a novel isolated polypeptide (NOXV). A
CC polypeptide of the invention has cytostatic, immunomodulator,
CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
CC antipneumatic activity, and may have a use in gene therapy, and as a

CC vaccine. The polypeptides are encoded by NOXV polynucleotides comprising
CC any of the 303 fully defined nucleotide sequences given in the
CC specification. The polypeptide is useful in the manufacture of a
CC medicament for treating a syndrome associated with a human disease. The
CC polypeptide, polynucleotide and antibody are useful in diagnosing,
CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,
CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
CC diseases, metabolic syndrome X or dyslipidemia. The nucleic acids are
CC further used as hybridisation probes, in chromosome mapping, tissue
CC typing, preventive medicine, and pharmacogenomics. The present sequence
CC encodes a NOXV polypeptide of the invention.
XX

SQ Sequence 12900 BP; 2893 A; 3775 C; 3410 G; 2822 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	12900
Score:	15519.00	Matches:	2887
Percent Similarity:	93.44%	Conservative:	21
Best local Similarity:	92.77%	Mismatches:	46
Query Match:	91.37%	Indels:	160
DB:	12	Gaps:	12

US-10-016-248-2 (1-3104) x ADH71145 (1-12900)

QY 2 AlaGlyAlaProProAlaLeuLeuProCysSerLeuLeu---SerAspCys--- 19
DB 731 TCTGCTTACCGAGCCAGCTCCACCCCGCTTACAGCAGCAAAACCTGCGCAGC 790
QY 20 CysAlaSerAsnGlnArgHisSerValGlyValGly-----ProSerGluLeuVal 36
DB 791 TGCACCTTCACATCGAGTGCAGCAACACCGCAGCGCGGATTCAGTCCCAATATCA 849
QY 37 LysLeuGlnIleGluLeuLysSerArgGlyValLysLeuMetProSerLysAspAsnSer 56
DB 850 AAGAGCAAAATGAGTGAAGTCTCGAGGTGAAGTGAATGCCACCAAGACAAACAGC 909
QY 57 GlnLysThrSerValLeuThrGlnValGlyValSerGlnLysHisAsnMetCysProAsp 76
DB 910 CAGAGACGCTGCTGGTGAATCACTCAGGTGTGTGTGCCAAGACATATATGTGCCAGC 969
QY 77 ProGlyLysProGluArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerVal 96
DB 970 CTTGGCATACCCGAAAGGGCCAAAGACTAGGCTCGGATTTACAGTTAGGATCCAGCGTC 1029
QY 97 GlnPheThrCysAsnGlnGlyLysArgLeuGlnGlySerLysArgLysIleThrCysMetLys 116
DB 1030 CAGTTCACTGCAACGAGGGCTATGACTCAAGGGTCCAAAGGGATCAGCTGTATGAAA 1089
QY 117 ValSerAspMetPheAlaIleTyrSerAspHisArgProValCysArgAlaArgMetCys 136
DB 1090 GTGAGCGACATGTTTGGCGCTGAGGCGACACAGGCCATCTGCCAGCCCGCATGTGT 1149
QY 137 AspAlaHisLeuArgGlyProSerGlyLysIleThrSerProAsnPheProIleGlnTyr 156
DB 1150 GATGCCACCTTCGAGGCCCTCGGGCATCATCACTCCCATTTCCCATTCAGTAT 1209
QY 157 AspAsnAsnAlaHisCysValIleTyrIleIleThrAlaLeuAsnProSerLysValIleLys 176
DB 1210 GACAAACAATGACACCTGTGTGTGATCATCAGCATCAACCCCTCCAGGAATATCAAG 1269
QY 177 LeuAlaPheGluGluPheAspLeuGluArgGlyTyrAspThrLeuThrValGlyAspGly 196
DB 1270 CTGCGCTTTAGAGAGTTGATTTGGAGAGGGGCTATGACCCCTGACGGTGGTATGCT 1329
QY 197 GlyGlnAspGlyAspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSer 216
DB 1330 GGTCAAGATGGGGCCGAGAGACAGTTCTTACATG----- 1365
QY 217 ProHisThrProGlySerArgGlyLeuProGluSerMetSerGlyAspIleTyrAspGlnLys 236
DB 1366 -----CTGACAGATACAGGTCCTCCGAGATCATGTGCAGCAAT---CATCAATG 1416
QY 237 TrpThrValLeuGluIleCysArgAspIleSerSerSerAspAlaArgSerGlySerVal 256

```

Db      1417 TGGCTCTCTTCAG-----ACTGATGC---AGTGGCAGT---- 1449
Qy      257 ArglySerProlySerAspAlaValIGluLeuValAlaProGlyThrGluIleGlu 276
Db      1450 -----TCCCTGGGATTCAGAGCTCTTATGAAAGAGATCCAG 1485
Qy      277 GInglySerCysGlyAspProGlyIleProAlaTyrGlyArgArgGluGlySerArgPhe 296
Db      1486 CAGGCGAGTTGGCGTGAACCTTGACATCTGATATGGCCGAGGGAAGCTCCCGGTTT 1545
Qy      297 HisHISGlyAspThrLeuLysPheGluCysGlnProAlaPheGluLeuValGlyGlnLys 316
Db      1546 CGCCACGGTGACACTCAAGTTTGTAGTCCAGCCCGCTTGAGCTGTGGACAGAG 1605
Qy      317 AlaIleThrCysGlnLysAsnAsnGlnIlePheSerAlaLysLysProGlyCysValPheSer 336
Db      1606 GCAATCATGCTCCAAAGATACCAATGATGGCTGAAGAGCCAGGCTGGTGTTC 1665
Qy      337 CysPhePheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAsp 356
Db      1666 TGCTTCTTCAACTTCACAGCCCGCTGGGGTGTTCGTCTCCCACTACCCAGAGAC 1725
Qy      357 TyrGlyAsnHisIleHisCysValIThrLeuIleLeuAlaArgProGluSerArgIleHis 376
Db      1726 TATGGCAACACCTTCACCTGTGTCTGCTCATCTGGCCAGGCTGAGAGCCGATCAC 1785
Qy      377 LeuAlaPheAsnAspIleAspValGluProGlnPheAspPheLeuValIleLysAsnGly 396
Db      1786 CTGGCTTCAACGACATTCGATGAGTGGAGCTCATGTTATTTCCGATCATCAAGAGTGG 1845
Qy      397 AlaThrAlaGluAlaProValLeuGlyThrPheSerGlyAsnGlnLeuProSerSerIle 416
Db      1846 GCCACCGCGAGGCGCGCTCTGGGACCTTCAGAGAAACGAGCTTCCCTCCATC 1905
Qy      417 ThrSerSerGlyHisValAlaArgLeuGluPheGlnIleThrAspHisSerThrGlyLysArg 436
Db      1906 ACAAGCAGTGGCCACGTGGCCCGCTCTGAGTTCCAGACTACCACTCCAGAGGAAGG 1965
Qy      437 GlyPheAsnIleThrPheThrThrPheArgHisAsnGluCysProAspProGlyValPro 456
Db      1966 GGCCTTCAACATCATCTTTACAGTGAAGTCTCCCAAGAGGCCGATCTGGCCATTCCA 2025
Qy      457 ValAsnGlyLysAspArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCys 476
Db      2026 GTAAATGCAAAACGGTTTGGAGACAGCTTCAGCTGGGACGCTCCATCTCTCTCTGT 2085
Qy      477 AspGluGlyPheLeuGlyThrGlnGlySerGluThrIleThrCysValLeuLysGluGly 496
Db      2086 GATGAAGGCTTCTCTGGGACTCAGAGGCTCAGAACCATCTGCTGCTTGAAGAGGAGC 2145
Qy      497 SerValValIThrAsnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisLeuThr 516
Db      2146 AGCGTGTCTGGAACAGCGCTGTGCTGCGGTGAAGCTCCCTGTGTGTGACTGACT 2205
Qy      517 SerProSerGlyThrIleLeuSerProGlyTyrProGlyPheThrLysAspAlaLeuSer 536
Db      2206 TCCCCAGCGGACCAATCTCTCTCCGGGCTGGCTGGGCTTCTACAAAGATCCCTTGAC 2265
Qy      537 CysAlaITrValIleGluAlaGlnProGlyTyrProIleLysIleThrPheAspArgPhe 556
Db      2266 TGTGCTGGGTGATGAGGCCAGCGCTACCCATCAAAATCATCTTCCACAGGTTT 2325
Qy      557 LysThrGluValAsnTyrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaPro 576
Db      2326 AAAACGAGGTCAACATGACACCTCGAAGATACGCGATGGCGGACTTACACGCGCC 2385
Qy      577 LeuIleGlyValTyrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAsnTyr 596
Db      2386 TTGATGGGGTTTACACGAGACCCAGGTTCCCACTTCTCATCGACCAACCACTAC 2445
Qy      597 LeuTyrLeuLeuPheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgTyr 616

```

```

Db      2446 CTCTACCTCTCTTCTCTACCAAGAGTCACTCGACATGGGCTTCCAGCTCCGCTAT 2505
Qy      617 GluThrIleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGln 636
Db      2506 GAGGCTATAACTGACAGTCAAGCACCTGTCTGGATCCAGAAATCCCATTAATGACAG 2565
Qy      637 ArgHisGlyAsnAspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyr 656
Db      2566 GGTCAATGGAATGATCTTACGTGGCGGCTGGTGAACCTTCAGCTGTACCTCGGCTAC 2625
Qy      657 ThrLeuSerAspGlyGluProLeuGluCysGluProAsnPheGlnITrSerArgAlaLeu 676
Db      2626 ACATTAAAGTACGGGAGGCTCTGAGTGTGAGCCCACTTCAGTGAAGCCGGGCTTG 2685
Qy      677 ProSerCysGluAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSer 696
Db      2686 CCCAGTTTGAAAGCTCTCTGTGGCTTCATTCAAGGCTCCAGTGGGACCATCTTGTCG 2745
Qy      697 ProGlyPheProAspPheTyrProAsnLeuAsnGlyThrIleGluThrSer 716
Db      2746 CAGGTTTCCCTGACTTCTACCCCAACACTTGAACCTGCATCTGATTAACAAACATCT 2805
Qy      717 HisGlyLysGlyValPhePheThrPheHisIleGluLysSerGlyHisAspTyr 736
Db      2806 CATGGCAAGGAGTGTCTTCACTTCCACACCTTCCAGCTGGAAAGTGGCATGACTAC 2865
Qy      737 LeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArg 756
Db      2866 CTCTCATCATCAGAAAGGAGGAGCTTCAACGAGCCCTGAGAGGAGCTAATGSAATCTCG 2925
Qy      757 LeuProAlaProIleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIle 776
Db      2926 CTGCAGCTCCCATAGCGCTGGGCTCTATGGCACTTCACTGCCAGAGTCCGCTTCACTC 2985
Qy      777 SerAspPheSerMetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGlu 796
Db      2986 TCTGATTTCTCATGTATGAAAGATTCAACTCATCACTTCTCAGAGTGAAGCTGGAG 3045
Qy      797 ProCysGluGluProGluValProAlaTyrSerIleArgGlyGlyLeuGlnPheGlyVal 816
Db      3046 CCTCTGAGAGCCCGAGAGTCCAGCTCAAGCATCCGGAAGGGCTTGCAGTTTGGCGTGG 3105
Qy      817 GlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIle 836
Db      3106 GCGGACACTTGACCTTCTCTGCTTCCCGGATACCGTCTGAGAGGACCGCCGCACTC 3165
Qy      837 ThrCysLeuGlyGlyValArgArgLeuITrSerSerProLeuProArgCysValAlaGlu 856
Db      3166 ACGTGCCTGGGGGACAGCGCGCTGTGAGCTCGCTGCGCCAGAGTGTGTGTGAG 3225
Qy      857 CysGlyAsnSerValIThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsn 876
Db      3226 TGTGGAAITTCAGTCAAGGACATCAAGGTAATTTGCTGTCGCCCACTTCCGTGAAC 3285
Qy      877 TyrAsnAsnAsnHisGlyCysIleTyrSerIleGlnIThrGlnProGlyLysGlyIleGln 896
Db      3286 TACATTAACATCATGATCATCTCATCTCCATCAAGCCAGCAGGAGGAAGGAAATTCAG 3345
Qy      897 LeuLysAlaArgAlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsn 916
Db      3346 CTGAAGCGCAGGGATTCGAACCTTCGGAAGAGATGTCCTCAAGTTTATGTATGGCAAC 3405
Qy      917 AsnAsnSerAlaArgLeuLeuGlyValPheSerHisSerGluLeuMetCysValITrLeu 936
Db      3406 AACAATCCCGCCGTTTGTGGAGATTTTATAGCATTTTGAAGTATGGGGGATCTTGG 3465
Qy      937 AsnSerThrSerSerSerLeuTyrLeuAspPheIleThrAspAlaGluAsnITrSerLys 956
Db      3466 AACAGCACATCCAGAGCTGTGTGTGATTTTCATCATGATCTGAAAAACACCAACAAG 3525
Qy      957 GlyPheGluLeuHisIlePheSerSerPheGluLeuIleLysCysGluLysAspProGlyTyrPro 976
Db      3526 GGCCTTGAATCTGACATTTTCCAGTTTGAATCTCATCAAAATGTAGAGACCCAGAACCCCC 3585

```

Qy	977	LySPheGlyTYrLYsValHI:ASpGluGlyHISpHeAlaGlySerSerValSerPheSer	996
Db	3586	AAgTTGGCTACAAAGTTCAATGATGAAAGCTATTTGCAAGGAGCTCCGTGTCTTCCACG	3645
Qy	997	CysAspProGlyTYrSerLeuArgLYSerGluLeuLeuCYLeuSerGlyGluArg	1016
Db	3646	TGTGACCTTGATACAGCTCGGGGGTAGTGGAGGCTGTGTCTGATGAGAGAGCGC	3705
Qy	1017	ArgThrTrpAspArgProLeuProThrCYsValAlaGluCYsGlyGlyThValArgGly	1036
Db	3706	CGGACCTGGAGCCGGCTCTGCCACCTGTGTCCGAGTGGAGGACGTAGAGGA	3765
Qy	1037	GluValSerGlyGluValLeuSerProGlyTYrProAlaProTYrGluHI:ASnLeuAsn	1056
Db	3766	GAGGTGTGGGGGAGGTCTGTCTACCCGGGTATCCAGCTCCCTTTGAACAATCTCAAC	3825
Qy	1057	CysIleTPThriIleGluAlaGlyCYsThrIleGlyLeuHI:spHeLeuValPhe	1076
Db	3826	TGCATCTTGACCATCGAAGCAGAGCCGGCTGCACCATTTGGGCTTCACTTCTGTGTTT	3885
Qy	1077	AspThrGluGluValHISpValLeuArgIleTPAspGlyProValGluSerGlyVal	1096
Db	3886	GACACAGAGAGGTTCAAGAGCTGTCTGCGCATCTGGGATGGGCTGTGGAGAGCGGGGTT	3945
Qy	1097	LeuLeuLYsGluLeuSerGlyProAlaLeuProLYsAspLeuHI:AspThrPheAsnSer	1116
Db	3946	CTGCTGAAGAGCTGAGTGGCCCGGCTGTCCCAAGACCTGCATAGCACTTCAACTCG	4005
Qy	1117	ValValLeuGlnPheSerThrAspPhePheThrSerLYsGlnGlyPheAlaIleGlnPhe	1136
Db	4006	GTCGCTCGCAGTTGACGACATGCTTTCACAGCAAGAGGGCTTGGCATTCATTT	4065
Qy	1137	SerValSerThrAlaThrSerCYsAsnAspProGlyIleProGlnAsnGlySerArgSer	1156
Db	4066	TCAGGTTCCACAGAACCTCTCTGCATACCTCTGGATCCCGAGAAATGGAGTCGGAGT	4125
Qy	1157	GlyAspSerTPGluAlaGlyAspSerThrValPheGlnCYsAspProGlyTYrAlaLeu	1176
Db	4126	GGTACACTGGAGAGCCGGGAGCTCCACAGGTTCACAGTGAACCTCGGTACAGCGGTG	4185
Qy	1177	GlnIleSerAlaGluIleSerCYsValLYsIleGluAsnArgPhePheTPRGLnProSer	1196
Db	4186	CAGGAGATGCAGAGATCAGCTGTGTGAAGATCGAAGACAGTTCTTGGCAACCCAGC	4245
Qy	1197	ProProThrCYsIleAlaProCYsGlyGlyAspLeuThrGlyProSerGlyValIleLeu	1216
Db	4246	CCGCCAATGATCGCTCCCTGGCGGGGAGACCTGCACAGACATCTGGAGTCATCTCTC	4305
Qy	1217	SerProAsnTYrProGluProTYrProProGlyLYsGluCYsAspTPRlyValThVal	1236
Db	4306	TCACCAAAATTAACCAAGAACCTTACCCGCAAGAGAGTGTGATGAAAGTAAACCGTCTC	4365
Qy	1237	SerProAspTYrValIleAlaLeuVal---PheAsnIlePheAsnLeuGluProGlyTYr	1255
Db	4366	TCACCAAGACTACGTATCGCCCTGTGTATGTTCCCAAGCTTTAACTGGAGACCTGGCTAT	4425
Qy	1256	AspPheLeuHI:AlleTYrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTYr	1275
Db	4426	GACTTCCCTCCATCTACGAGGAGCGGACTCTCTCAGCCCTCTCATAGGAAGCTTCTAT	4485
Qy	1276	GlySerGlnLeuProGlyArgIleGluSerSerAsnSerLeuPheLeuAlaPheArg	1295
Db	4486	GGCTCCCAAGCTCCCAAGGCGCATTTGAAGCAGCAACACCTTCTCTGCTTCCGCGC	4545
Qy	1296	SerAspAlaSerValSerAsnAlaGlyPheValIleAspTYrThrGluAsnProArgGlu	1315
Db	4546	AGCATGTGATGTGAGCAATGCTGGCTTCTCATTTGACTTCCAGAAAAACCCCGGGAG	4605
Qy	1316	SerCYsPheAspProGlySerIleLYsAsnGlyThArgValGlySerAspLeuLYsLeu	1335
Db	4606	TCATGTTTGTATCTGTGTTTCATCAAGAACCGGAGCGGCTGGAGCTGCAACTGAAGCTG	4665
Qy	1336	GlySerSerValThrTYrTYrCYsHI:GlyGlyTYrGluValGluGlyThrSerThrLeu	1355
Db	4666	GGCTCTCCGTCACTCACTCACTGCGACAGGGGGCTACGAAGTTGAGGGCACTGCACCTG	4725
Qy	1356	SerCysIleLeuGlyProAspGlyLYsProValTPRAsnAspProArgProValCYsThr	1375
Db	4726	AGTGCATCTCGGGGGCTGATGGGAAGCCGTTGGAAACATCCCGGCAAGTCTGCACA	4785
Qy	1376	AlaProCYsGlyGlyGlnTYrValGlySerAspGlyValValLeuSerProAsnTYrPro	1395
Db	4786	GGCCCTGTGGGGGACAGTATGTGGGTGTGAGCGAATGTGTCTTCCCACTACACCC	4845
Qy	1396	GlnAsnTYrThrSerGlyGlnIleCYsLeuTYrPheValThrValProLYsAspTYr---	1414
Db	4846	CAGAACTACACCAATGAGACAGATCTGTTGATTTTGTACTGTGCCCAAGACTAAGT	4905
Qy	1415	ValValPheGlyGlnPheAlaPhePheHI:sthrAlaLeuAsnAspValAlaGluValHIS	1434
Db	4906	GTTGTTGTTGGCCAGTTGCGCTTCTTTCACACGGCCCTCAACGACGTGTGAGGTTCAC	4965
Qy	1435	AspGlyHI:SerGlnHI:SerArgLeuLeuSerSerLeuSerGlySerHI:sthr---	1453
Db	4966	GACGGCCACAGCCAGCACTGGCGCTCTCAGCTCCCTTCGGGCTCCCATACAGGTGA	5025
Qy	1454	GlySerLeuProLeuAlaThrSerAsnGlnValLeuIleLYsPheSerAlaLYsGlyLeu	1473
Db	5026	GAATCACTGCTTGGCCACTTCCATCAAGTTCTCTTAAGTTACAGCCCAAGAGCTTC	5085
Qy	1474	AlaProAlaArgGlyPheHI:spHeValTYrGlnAlaValProArgTPRserAlaThrGln	1493
Db	5086	GCACACAGCAGAGGCTTCCACTTGTCTACCAAGCGGTTCTCGAACCAAGCAGCAG	5145
Qy	1494	CysSerSerValProGluProArgTYrGlyLYsArgLeuGlySerAspPheSerValGly	1513
Db	5146	TGACGCTGTGTGGCGAAACCCCTATGGCAAGAGCTGGGAGTGACTTCTGTGTGGG	5205
Qy	1514	AlaIleValArgPheGluCYsAsnSerGlyTYrAlaLeuGlnGlySerProGluIleGlu	1533
Db	5206	GCCATGCTCGGCTTCCAGATCCAACTCCGCTTAGTCCCTGACAGGGGTGCGAAGATGAG	5265
Qy	1534	CysLeuProValProGlyAlaLeuAlaGlnTPRAsnValSerAlaProThrCYsValVal	1553
Db	5266	TGCTCCCTGTGCTGGGGCTTGGCCCAATGGAATGTCCAGCGCCACAGTGTGTGTG	5325
Qy	1554	ProCYsGlyGlyAsnLeuTPRgluArgArgGlyThrIleLeuSerProGlyPheProGlu	1573
Db	5326	CCGTGTGAGGCAACCTCAAGAGCGGAGGAGCCATCTGTGCTCCCTGCTTCCAGAG	5385
Qy	1574	ProTYrLeuAsnSerLeuAsnCYsValTPRlyIleValValProGluGluAlaGlyIle	1593
Db	5386	CCGTACTTCAACAGCTCACTGTGTGTGGAAGTGTGTGCCCCGAAAGGCTGGCATC	5445
Qy	1594	GlnIleGlnValValSerPheValThrGluGlnAsnTPRAspSerLeuGluValPheAsp	1613
Db	5446	CAGATCCAAATTTGCAATTTGTGACAGACAGAACTGGGACTCGGTGAAGATTTGAT	5505
Qy	1614	GlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThValProAlaLeu	1633
Db	5506	GGTGCAGATTAACCTGAACCATGTGGGAGTTTCCAGAAACAACCGTGGCTGCTT	5565
Qy	1634	LeuAsnSerThrSerAsnGlnLeuTYrLeuHI:spHeTYrSerAspIleSerValSerAla	1653
Db	5566	CTGAACAGACCTTCAACAGCTCTACCTTCACTTCTACCAATATCAACGATTCGCA	5625
Qy	1654	AlaGlyPheHI:AlleuGluTYrLYsThValAlaGlyLeuSerSerCYsProGluProAlaVal	1673
Db	5626	GCTGGCTTCACTTGGAGTACAAGACGTGTGGCTGAGCAGTGTCCGAAACCTGGCTG	5685
Qy	1674	ProSerAsnGlyValLYsThrGlyGluArgTYrLeuValAsnAspValValSerPheGln	1693
Db	5686	CCCAAGTAACGGGGTGAAGACTGGAGCGGCTACTTGTGTGATGATGTGTGTCTTCCAG	5745
Qy	1694	CysGluProGlyTYrAlaLeuGlnGlyHI:AlaHI:AlleSerCYsMetProGlyThrVal	1713

```
Db 5746 TGTGAGCCGGGATATTCCTCCAGGGCCAGCCCACTTCCTGATGCGGAAACAGTG 5805
Qy 1714 ATGAGGTTPAANTYrProBProleuCySIIlealaginCySgIyGlyThrValGluGlu 1733
Db 5806 CCGGCATGTGACATCTCCCTCCACTGTATTTGACAGTGTGGGGAAACAGTGGAGAG 5865
Qy 1734 MetGluGluValIIleuSerProGlyPheProGlyAsnTYrProSerAsnMetAspCys 1753
Db 5866 ATGGAGGGGGTATCTTGAGCCCCGGCTTCCAGGCAACTACCCAGTAACATGAGACTGC 5925
Qy 1754 SerTPlySIIlealIleuProValGlyPheGlyAlaniSIIleGlnPheLeuAsnPhSer 1773
Db 5926 TCCGTGAAAATAGCACTGGCCGCTGGAGCTTGAGACTCAATCCAGTTCTTAATCTTCC 5985
Qy 1774 ThrGluProAsnHIIAspTYrIIleGluIIleArganGlyProTYrGluThrSerArgMet 1793
Db 5986 ACCGAGCCCAACCAAGACTTACATAGAAATCCGGAATGGCCCCCTATGAGACCAAGCCGATG 6045
Qy 1794 MetGlyArgPheSerGlySerGluLeuProSerSerLeuLeuSerThrSerHIIleGluThr 1813
Db 6046 ATGGGAAGATTCAGTGGAAGGAGCTTCCAGCTCCCTCTCTCCAGCTCCAGACAGAC 6105
Qy 1814 ThrValTYrPheHIIAspHIIAspSerGlnAsnArgProGlyPheGlySLeuGluTYrGln 1833
Db 6106 ACCGTGATTTTCCACAGGACCACTCCAGAAATCGGCGAGATTCAAGCTGAGATACAG 6165
Qy 1834 AlaTYrGluLeuGlnGluCysProAspProGluProPheAlaAsnGlyIIleValArgGly 1853
Db 6166 GCCTATGAACTTCAAGAGTCCCAAGCCCAAGGCCCTTGGCAATGGCAATTTGAGGGGA 6225
Qy 1854 AlaGlyTYrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTYrGlnLeuThr 1873
Db 6226 GCTGGCTACAAAGTGGGACATCAGTGACCTTGAGTGGCTCCGGGGTATCAATTGACT 6285
Qy 1874 GlyHISProValLeuThrCysGlnHIIleGlyThrAsnArganTYrAspHIIleProleuPro 1893
Db 6286 GGCACCTCTGCTCAAGTGTCAACATGGACCAACCGGAATCGGAGCCACCCCTGGCC 6345
Qy 1894 LysCysGluValProCysGlyGlyAsnIIleThrSerSeranGlyThrValTYrSerPro 1913
Db 6346 AAGGTGAGAGTCCCTTGGGGGGAACATCATCTTCCACAGGCAT -GTGTCCTCCCG 6404
Qy 1914 GlyHerProSerProTYrSerSerSerGlnAspCysValTYrPheIIleThrValProIIle 1933
Db 6405 GGGTTCCTTACCCGCTCAAGCTCCAGACTGTGTGCTGCTGATACACGTCGACAT 6464
Qy 1934 GlyHISGlyValArgLeuAsnLeuSerLeuGlnThrGluProSerGlyAspPheIIle 1953
Db 6465 GGCATGCGCTCCGCTCAACCTCAAGCTGTGAGACAGAGCCCTGAGATTTTCATC 6524
Qy 1954 ThrIIleTPAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMet 1973
Db 6525 ACCATCTGGGATGGGCACAGCAACAGCACAGCGGTGGGCTTCCACCCGAGACAG 6584
Qy 1974 AlaIleValIleValGlnSerSerSerAsnGlnValLeuLeuLysPheHIIleArgAspAla 1993
Db 6585 GCCAAGAAACAGTGCAGAGTTCATCAACAGAGTCTCTCTCAAGTTCCACCGTATGCA 6644
Qy 1994 AlaThrGlyGlyIIlePheAlaIIleAlaPheSerAlaTYrProLeuThrLysCysProPro 2013
Db 6645 GCCACAGGGGGGATCTTCCGCAATGCTTCTCCGCTTATCCACTCAACCAATGGCCCTCT 6704
Qy 2014 ProThrIIleuProAsnAlaGluValValThrGluAsnGluGluPheAsnIIleGlyAsp 2033
Db 6705 CCCACATCTCCCAACCGCGAAGTCGTACAGAGAAATGAAGAAATTCATATAGGTGAC 6764
Qy 2034 IIleValArgTYrArgCysLeuProGlyPheThrLeuValGlyAsnGluIIleuThrCys 2053
Db 6765 ATGTGACGTACAGAGTCTCTCCGCTTATCCATTAAGTGGGAAATTCATGACCTGC 6824
Qy 2054 LysLeuGlyThrTYrLeuGlnPheGluGlyProProProIIleCysGluValHIIleCysPro 2073
Db 6825 AAACCTGGAACCTTACCTGAGATTGAAAGCAACCCCGATATGTGAAGTGCATGTCCA 6884
Qy 2074 ThrAsnGluLeuLeuThrAspSerThrGlyValIIleLeuSerGlnSerTYrProGlySer 2093
Db 6885 ACNAATGAGCTTTGACAGACTTCCACAGCGAGATCTTAAGCAGAGCTTACCTGGAAAC 6944
Qy 2094 TYrProGlnPheGlnThrCysSerTPleuValArgValGluProAspTYrAsnIIleSer 2113
Db 6945 TATCCCAATTCACAGACTGTCTTGCTGGTGTGAGATGGAGCCGAGCTATAACTTCC 7004
Qy 2114 LeuThrValGluTYrPheLeuSerGluLysGlnTYrAspGluPheGluIIlePheAspGly 2133
Db 7005 CTCACAGTGAAGTACTCTTCAAGCAAGAAATGAGATTTGAGATTTTATGATGT 7064
Qy 2134 ProSerGlyGlnSerProLeuLeuLysAlaLeuSerGlyAsnTYrSerAlaProleuIIle 2153
Db 7065 CCATCAGACAGAGTCTCTGCTGAAGGCCCTCAAGGAAATTAATCACTCCCTGAT 7124
Qy 2154 ValThrSerSerSerAsnSerValTYrLeuArgTYrPheSerAspHIIleAlaTYrAsnArg 2173
Db 7125 GTCCACAGCTCAAGCAACTGTGTACTGTGCTGTGATCATGTACAGCCCTACATCGG 7184
Qy 2174 LysGlyPheLysIIleArgTYrSerAlaProTYrCysSerLeuProArgAlaProleuHIS 2193
Db 7185 AAGGCTTCAGATCCGCTATTCAGCCCTTATCTGACACTGTGCCAGGGCTTCATCTCAT 7244
Qy 2194 GlyPheIIleuGlyGlnThrSerThrGlnProGlyGlySerIIleHISpHeGlyCysAsn 2213
Db 7245 GCCTTCATCTTACGCGCAGACAGCACCCAGCCGGGCTCATCTGTTGGCTGCAC 7304
Qy 2214 AlaGlyTYrArgLeuValGlyHISerMetAlaIIleCysThrArgHISpProGlnGlyTYr 2233
Db 7305 GCGGCTTACCGCTGTGTGGACACAGCAATGGCCATGTGTACCCGACACCCCAAGGGCTAC 7364
Qy 2234 HISLeuTPSerGluValIIleProleuCysGlnAlaLeuSerCysGlyLeuProGluAla 2253
Db 7365 CACTGTGAGAGGAGCATCTCTCTGTCAACTCTTCTTGAGCTTCTGAGGCC 7424
Qy 2254 ProLysAsnGlyMetValPheGlyLysGluTYrThrValGlyThrLysAlaValTYrSer 2273
Db 7425 CCCAAGATGGAAAGTGTGTTGGCAAGAGTACACAGTGGGAAACCAAGCCCATGTACAGC 7484
Qy 2274 CysSerGluGlyTYrHISLeuGlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThr 2293
Db 7485 TGCACTGAGGCTTACCACTTCCAGCGAGCGCTAGGCACTGAGAGTGTCTGGACACA 7544
Qy 2294 GlyLeuTPSerAsnArganValProProGlnCysValProValThrCysProAspVal 2313
Db 7545 GGCCTATGGAGCAACCCCAATGTCCACACAGTGTCTCCCTGTGACTTGTCCGATGTC 7604
Qy 2314 SerSerIIleSerValGluHISGlyArgTYrArgLeuIIlePheGluThrGlnTYrGlnPhe 2333
Db 7605 AGTACATCAAGCGTGGAGCAATGGCGGAGGCTTATCTTTGAGACACAGTACAGTTC 7664
Qy 2334 GlnAlaGlnLeuMetLeuIIleCysAspProGlyTYrTYrTYrThrGlyGlnArgValIIle 2353
Db 7665 CAGGCCAGCTGATGCTCATCTGACCTCGCTACACTACATACCTGCCCAAGGGTATC 7724
Qy 2354 ArgCysGlnAlaAsnGlyLysTYrPheSerLeuGlyAspSerThrProThrCysArgIIle 2373
Db 7725 CGCTGTAGGCCAATGGCAATGAGAGCTTGGGAGACTTACGCGCCACTTCCGAATCATC 7784
Qy 2374 SerCysGlyGluLeuProIIleProProAsnGlyHISArgIIleGlyThrLeuSerValTYr 2393
Db 7785 TCCGTGTGAGAGCTCCGATTTCCCCCAATGGCCACCGCATCGGAACACTGTGTGCTAC 7844
Qy 2394 GlyAlaThrAlaIIlePheSerCysAsnSerGlyTYrThrLeuValGlySerArgValArg 2413
Db 7845 GGGGCAACAGCACTTCTCTCTCAATTCGGAATACACTGTGTGGCTCCAGGGTGGCG 7904
Qy 2414 GluCysMetAlaAsnGlyLeuTPSerGlySerGluValArgCysLeuAlaGlyHISCys 2433
Db 7905 GAGTGAAGGCCAATGGCTGTGAGTGGCTGTGAAGTCCGCTTCTGTGACACTGT 7964
```

OY	2434	GIYTHRRPROGILUPROILIVAlAsnLIYHILLeAsnGIYGIUAsnTYSerTYrARGLY	2453
Db	7965	GGAGTCTCTGAGCCCATTTGTCAACGACACATCAATAGGAGAACTTAACGCTACCGGGGC	8024
OY	2454	SeValValIYrGInCyAsnAlaGILyPheArGLeuIleGILyMetSerValArgIleCys	2473
Db	8025	AGTGGGTGATACCAATGCATGCAATGCTGGCTTCGGCTGATCGGAGATCTGTGCGATCTGC	8084
OY	2474	GIInGInARPHLHITrPSerGIYLySThrPrORheCysValProIleThrCysGLYHIS	2493
Db	8085	CAGCGAGATCATCACTGGTCGGGGCAAGACCCCTTCTGTGTGTC	8126
OY	2494	ProGIYAsnProValAsnGILyLeuThrGInGILyAsnGInPheAsnLeuAsnAlaVal	2513
Db	8126	-----	8126
OY	2514	LYSPheValCyAsnAsnProGIYTYrMetAlaGILyAlaAlaArgSerGInCysLeuAla	2533
Db	8126	-----	8126
OY	2534	SerGIYGIInTrPSerAspMetLeuDrOThrCyARGLIleLeAsnCysThrAspProGIY	2553
Db	8126	-----	8126
OY	2554	HisGInGILyAsnSerValArgGInValHisAlaSerGIYProHILeARPhSerPheGLY	2573
Db	8126	-----	8126
OY	2574	ThrThrValSerTYrARGCysAsnHILeGILyPheTYrLeuLeuGILyThrProValLeuSer	2593
Db	8126	-----	8126
OY	2594	CysGInGILyAsnGILyThrTrpAspArgProARpProGInCysLeuLeuValSerCysGLY	2613
Db	8127	-----CTGGTCTCCGTGGCC	8141
OY	2614	HisProGIYSerProProHILeSerGInMetSerGIYAspSerTYrThrValGILyAlaVal	2633
Db	8142	CATCGGGGCTCCCGCCTCACTCCCAAGATGCTGGAGACAGTTAACTGTGGAGCACTG	8201
OY	2634	ValARGLYrSerCysIleGILyAsnArgThrLeuValGILyAsnSerThrArgMetCysGLY	2653
Db	8202	GTGGCGTACAGCTGCATCGGACAGCGTACTCGTGGGGAAACAGACCCGATGTGGGG	8261
OY	2654	LeuAspGILyHILeTrpThrGILySerLeuProHILeCysSerGIYThrSerValGILyAlaCys	2673
Db	8262	CTGGATGGACACTGAGACTGGCTCCCTCCCTCACTGCTCAGGAACCAAGCGTGGAGTTTGC	8321
OY	2674	GIYAspProGIYIleProAlaHILeGILyIleArgGILyGIYAspSerPheAspProGIYThr	2693
Db	8322	GGTGAACCTGGGATCCCGGCTCATGGCATCCGTTGGGGAGACAGCTTTGAITCCAGGACCT	8381
OY	2694	ValMetArgPheSerCysGILyAlaGILyHILeValLeuArgGILySerSerGILyArgThrCys	2713
Db	8382	GTGATAGCCCTTCAAGCTGTGAAGACTGGCACAGCGTCCGGGGATCGTCAAGAGGCACTGT	8441
OY	2714	GIInAlaAsnGILySerTrpSerGILySerGInProGILyCysGILyValIleSerCysGILyAsn	2733
Db	8442	CAAGCCAATGGCTCTGTGAGCGCGCTCGAGCCTGAGTGTGAGTATCTTGTGGGAAC	8501
OY	2734	ProGIYThrProSerAsnAlaArgValAlaPheSerAspGILyLeuValPheSerSerSer	2753
Db	8502	CCTGGGATCTCAAGATATGCCGAGTTGTGTTCACTATGGCTCGTGTTCACAGCTCT	8561
OY	2754	IleValIYrGILyCyAsnArgGILyTYrTYrAlaThrGILyLeuLeuSerArgHILeCysSer	2773
Db	8562	ATCGCTCTATGAGTGGCGGAGAGAACTAACGCAAGCGCTGCTCAGCGCTCACTGCTCG	8621
OY	2774	ValAsnGILyThrTrpThrGILySerAspProGILyCysLeuValIleAsnCysGILyAspPro	2793
Db	8622	GTCATATGTAACCTGGACAGGCACTGATCCCTGAGTGCCTCTGCATTAACCTGTGTGACCTT	8681

QY	2754	GlyIleProAlaAsnGlyLeuAArgLeuGlyValAsnAspPheArgTyrAsnLysThrValThr	2813
Db	8662	GGAATTCAGCCAAATGGCTTCGGCTGGGCAATGACTTCAGGTCAACAAACCTGTACA	8741
QY	2814	TyrGlnCysValProGlyTyrMetMetGlnSerHisArgValSerValLeuSerCysThr	2833
Db	8742	TATCATGTGTCTCCTGGCTATATATATGAGTCAATAGATATCTGTGCTGAGCTGACCC	8801
QY	2834	LysAspArgThrTTPAsnGlyThrLysProValCysAlaLeuMetCysLysProPro	2853
Db	8802	AAGGACCGGACATGGAATGGAAACCAACCCGCTGCAAACTCTCATGTCCAAGCCACT	8861
QY	2854	ProLeuLeuProAsnGlyLysValValGlySerAspPheMetTPGlySerSerValThr	2873
Db	8862	CCGCTCATCCCAATGGGAAGTGTG3GGGTGTGACTTCATGTGGGGCTCAAGTGTACT	8921
QY	2874	TyrAlaCysLeuGlnGlyTyrGlnLeuSerLeuProAlaValPheThrCysGlnGlyAsn	2893
Db	8922	TATGCTCTGCTGGAGGGGTACCACTCTCCTGCGCGGGTGTTCACCTGTGAGGAAAT	8981
QY	2894	GlySerTPThrGlyGlnLeuProGlnCysPheProValPheCysGlyAspProGlyVal	2913
Db	8982	GGGTCTCTGACCGGAGAGCTGCTCAGAGTTTCTCCTGTGTTCTGGGGAGTCCGTGTCT	9041
QY	2914	ProSerArgGlyValArgArgGlyLysAspArgGlyPheSerTyrArgSerSerValSerPheSer	2933
Db	9042	CCGTCCCTGGGAGAGAGAGACCGAGGCTTCTCCACAGGTCATCTGCTCTCTCTTCTCC	9101
QY	2934	CysHisProProLeuValLeuValGlySerProArgArgPheCysGlnSerAspGlyThr	2953
Db	9102	TGCATCTCCCTCTGTGTCTGTGTGGCTCTTCACGCAAGTTTTCACGTAGATGGAGCA	9161
QY	2954	TTPSerGlyThrGlnProSerCysAlaAspProThrLeuThrThrCysAlaAspProGly	2973
Db	9162	TGAGGTGCAACCGACCGAGCTGATAGATCCGACCTGACACAGTGTGGGACCTGTGT	9221
QY	2974	ValProGlnPheGlyIleGlnAsnAsnSerGlnGlyTyrGlnValGlySerThrValLeu	2993
Db	9222	GTGCCACAGTTTGGGATACAGAACAAATTTCTCAGGGCTACACAGTTGGAAACACAGTCTTC	9281
QY	2994	PheArgCysGlnLysGlyTyrTyrLeuLeuGlnGlySerThrThrArgThrCysLeuProAsn	3013
Db	9282	TTCCGTGTCAAAAGGCTACCTGTCTTCAAGGCTCCACACACAGACTGCTCTCCCAAC	9341
QY	3014	LeuThrTTPSerGlyThrProProAspCysValProHisCysArgGlnProGluThr	3033
Db	9342	CTGACCTGGATGGAAACCCCACTGACTGTGTCCCCCAACACACTGACGACGACGAGAGCG	9401
QY	3034	ProThrHisAlaAsnValGlyValAlaLeuAspLeuProSerMetGlyTyrThrLeuIle-Th	3053
Db	9402	CCAACGCAATGCCAATGCTGGGGCCCTGGAATTTGGCCCTCCATGGGCTTCACAGCTCATCTAC	9461
QY	3053	rProAlaArgArgLysSerProSerArgValAlaProSerThrAlaProAlaArgArgMet	3073
Db	9462	TCTGTCCACGAGGGCTTCTCTCTCAAGGTGGCTCCAGACCGCACCTTCAAGCGCAT	9521
QY	3073	AlaAlaGlyValAlaSerArgProSerAlaTTPArgSerGlyProValGlyAspProse	3093
Db	9522	GGCAGCTGGACAGGCAAGCCGCCCATCTGCTGGAGGTCCGGCCCAATGGGAATCCATC	9581
QY	3093	rThrLeuProGlySerHisArgSerProLysPro 3104	
Db	9582	AACACTGCGCGGAGCCACCGCTACCCCAAGCCT 9615	
RESULT 9			
ABSE64376			
ID	ABSE64376	standard; DNA; 8010 BP.	
XX	ABSE64376;		
XX			
DT	15-NOV-2002	(first entry)	
XX			
DE	Human cub and sushi domain containing gene #2.		


```
QY 301 ThrLeuYsPheGluCySGlnProAlaPheGluLeuValGlyGlnLysAlaIleThrCys 320
DB 901 ACACCTCAAGTTGAGTGTCCACCGCCCTTTAGAGCTGTGGAGCAGAAAGGCATACATC 960
QY 321 GlnLysAnbAnGlnTPSerAlaLysLysProGlyCysValPheSerCysPhePheAsn 340
DB 961 CAAAGAATACCAATGGTGGCTTAAGAAAGCCAGGCTCGGTCTCTCGCTTCTTCAAC 1020
QY 341 PheThrSerProSerGlyValValLeuSerProAnTYrProGluAspTYrGlyLysHis 360
DB 1021 TTCACCAAGCCGCTGGGGTGTCTCTCTCCCACTACCAAGAGGACTAAGGCAACAC 1080
QY 361 LeuHisCysValTyrPheLleLeuAlaArgProGluSerArgIleHisLeuAlaPheAsn 380
DB 1081 CTCACAGTGTCTGAGCTCATCTGAGCCAGGCTGAGAGCCGCACTCACCTGGCTTCAAC 1140
QY 381 AspIleAspValGluProGlnPheAspPheLeuValIleLysAspGlyAlaThrAlaGlu 400
DB 1141 GACATTGACGTGAGGCTCAGTTGATTCTGTGTCATCAAGAGTGGGCCACCGCGAG 1200
QY 401 AlaProValLeuGlyLysPheSerGlyLysGlnLeuProSerSerIleThrSerSerGly 420
DB 1201 GCGCCCGCTCGGCGACCTTCTCAGAAACCAAGCTTCTCTCCATCACAAAGCAGTGGC 1260
QY 421 HisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLysArgGlyPheAsnIle 440
DB 1261 CACGTGGCCGCTCTGAGTTCAAGATTCAGACCACTCCACAGGAGAGGGGCTTCAACATC 1320
QY 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyLys 460
DB 1321 ACTTTTACCACTTCCGACCAACAGAGTGGCCGGATCTGGCGTTCCAGTAAGTGGCAA 1380
QY 461 ArgPheGluAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGluGlyPhe 480
DB 1381 CGGTTTGGGACACCTCCACCTGGCGGCGCTCCATCTCTCTGTAAGAGGGCTTC 1440
QY 481 LeuGlyThrGlnGlySerGluThrIleThrCysValLeuLysGluGlySerValValTyr 500
DB 1441 CTTCGAGCTCAGGCTCAGAACCACTACCTGCGCTCGAAGAGAGGGAGGCTGTCTGG 1500
QY 501 AsnSerAlaValLeuArgCysGluAlaProCysGlyGlyHisLeuThrSerProSerGly 520
DB 1501 AACAGCCCTGTGCTGGCGGTGAGGCTCCCTGTGTGTCTACCTGACCTGCCCAAGGCG 1560
QY 521 ThrIleLeuSerProGlyTyrProGlyPheTyrLysAspAlaLeuSerCysAlaTyrVal 540
DB 1561 ACCATCTCTCTCCGGGCTGGCTGGCTTCAAGAGATGCTTGAGCTGTGGGCTG 1620
QY 541 IleGluAlaGlnProGlyTyrProIleLysIleThrPheAspArgPheLysThrGluVal 560
DB 1621 ATTGAGGCCAGCAGGCTACCCATCAAAATCACTTCGACAAATTCAAAACCGAGGCTC 1680
QY 561 AsnTYrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaProLeuIleGlyVal 580
DB 1681 AACATATACACCTCGAAGTACGAGTGGGCGGACTTACTAGGCGCTTGATGGGGTT 1740
QY 581 TyrHisGlyThrGlnValProGlnPheLeuIleSerThrSerAnTYrLeuTyrLeuLeu 600
DB 1741 TACACCGGAGCCAGGTTCCCACTTCTCATCAGACACCACTACCTCACTCTCTC 1800
QY 601 PheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuArgTyrGluThrIleThr 620
DB 1801 TTCCTACCCAGCAAGATCACTCGACATCGGCTTCAGCTCCGCTATGAACTATATACA 1860
QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsn 640
DB 1861 CTGAGTCAAGCACTGTCTGATCCAGGAATCCCAATGAATGACACGCGTCAAGGAAT 1920
QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660
DB 1921 GACTTCAACGTGGCGGCTGTGACCTTCACTGTGACCTGGGCTACACATTAAGTAC 1980
QY 661 GlyGluProLeuGluCysGluProAsnPheGlnTPSerArgAlaLeuProSerCysGlu 680
DB 1981 GGGAGGCTCTGAGTGTGAGCCCACTTCCAGTGGAGCGGGGCTTGCCAGTTGTGA 2040
QY 681 AlaLeuCysGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
DB 2041 GCTCTGTGTGGCTTCATTCAAGGCTCCAGTGGAGCACTTGTGCGCAGGGTCCCT 2100
QY 701 AspPheTyrProAnbAnLeuAsnCysThrThrIleIleGluThrSerHisGlyLysGly 720
DB 2101 GACTTCAACCCCAACCACTTGAACTGCACTGTGATTATGAAATCTCATGGAAGGT 2160
QY 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTYrLeuLeuIleThr 740
DB 2161 GTGTTCTTCACTTCCACACTTCCACTGGAAGTGGCCATGACTCTCTCATACT 2220
QY 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
DB 2221 GAGAAAGGAGCTTCAACCCAGCCCTGAGGAGAGTAACTGGATCTCGCTGCGAGCTCC 2280
QY 761 IleSerAlaGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
DB 2281 ATCAGCGCTGGGCTCTATGGCACTTCACTGCGCAGGCTCGCTCATCTGATTTCTCC 2340
QY 781 MetSerTyrGluGlyPheAnbIleThrPheSerGluTyrAspLeuGluProCysGluGlu 800
DB 2341 ATGTCAATATAGAGATTCAACATCACTTCTCAGAGTACAGACTTGGAGCCCTGTGAAG 2400
QY 801 ProGluValProAlaTyrSerIleArgLysGlyLeuGlnPheGlyValGlyAspThrLeu 820
DB 2401 CCGAGGCTCCAGCTACAGATCCGGAAGGGCTTGAGTTGGCGTGGGAGCACTTGG 2460
QY 821 ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840
DB 2461 ACCTCTCTGCTTCCCGGGGTACCGTCTGAGGGCACCGCCGCACTCACTGCTGGG 2520
QY 841 GlyArgArgLeuTPSerSerProLeuProArgCysValAlaGluCysGlyAsnSer 860
DB 2521 GGCAGACGGCGCTGTGAGACTGCTCTGCAAGGTGTGTGTGTGTGGGAAATTC 2580
QY 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTYrAsnAsnAsn 880
DB 2581 GTCAAGAGCACTAGGCTACTTGTGTCTGCCCAACTTCTGTGAACTCAATACAT 2640
QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyLysGlyIleGlnLeuLysAlaArg 900
DB 2641 CATGAATGCACTTCACTCAGACCAAGGAGGAGGAAATTCAGCTGAAGAGCCAGG 2700
QY 901 AlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsnAnbAnSerAla 920
DB 2701 GCATTGAACTCTCCGAGAGAGATGCTCAAGGTTTATATGGAACAACAACCTCGCC 2760
QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAnbAnSerThr 940
DB 2761 CGTTTGTGGAGTTTATACCATTCGATATATAGGGGTGACTTTGAACAGACATCC 2820
QY 941 SerSerLeuTyrLeuAspPheIleThrAspAlaGluAnbAnThrSerLysGlyPheGluLeu 960
DB 2821 AGCAGTCTGTGGCTTGTATTCATCACTGAGTGAAGAACCAAGAGGCTTTGAAGCTG 2880
QY 961 HisPheSerSerPheGluLeuLysCysGluAspProGlyThrProLysPheGlyTyr 980
DB 2881 CACTTTCAGCTTGAATCATCAATGAATGAGAGCCAGAAACCCCAAGATTGGCTAC 2940
QY 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
DB 2941 AAGGTTCAATATAGAGTATTTTGAAGAGAGCTCGGTCTTCACTGCTGACCTCGA 3000
QY 1001 TyrSerLeuArgLysSerGluGluLeuLeuCysLeuSerGlyGluArgArgThrTyrAsp 1020
DB 3001 TACAGCTCTGGGGGTGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3060
QY 1021 ArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluValSerGly 1040
```


Db 3061 CGGCTCTGCGCCACCTGTCTCGCGAGTGTGAGGAGCAGTGAAGAGAGGTGTGGGG 3120
Qy 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleTyrThr 1060
Db 3121 CAGGTGCTGTCAACCCCGGATACAGCTCCCTATGAACAATCTCAATCGATCTGGACC 3180
Qy 1061 IleguaIagIuaIagIyCyrThrIlegIyLeuHisAspLeuValPheAspThrGlu 1080
Db 3181 ATCGAAGCAGAGGCGCGCTGCACCTTGGGCTTACCTTGTGTGTGGACACAGAGAG 3240
Qy 1081 ValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyValLeuLeuGlu 1100
Db 3241 GTTCACGACGTGCTCGCATCTGGAGTGGGCTGTGGAGAGGGGTTCTCTGAAGAG 3300
Qy 1101 LeuSerGlyProAlaLeuProIyAspLeuHisSerThrPheAsnSerValValLeuGln 1120
Db 3301 CTGATGGCGCCGCGCTCGCCACGAGACCTGCATACCTTCACTCGGTGTCTGACAG 3360
Qy 1121 PheSerThrAspPhePheThrSerIySGlnGlyPheAlaIlegIlnPheSerValSerThr 1140
Db 3361 TTCACACCTGACTTCTTCACGACGAGAGGGGCTTGGCATTTCAATTTTCAGTTCACA 3420
Qy 1141 AlaThrSerCysAsnAspProGlyTleProGlnAsnGlySerArgSerGlyAspSerTyr 1160
Db 3421 GCAACGTCTGTCAATGACCTCGGAGATCCGACAGATGGAGTCCGAGTGTGACAGTTGG 3480
Qy 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGlySerAla 1180
Db 3481 GAAGCCGCGACTCCACAGGTTCCTCACTGTACCTTGTGACGCGGTGACGGGAGAGTCA 3540
Qy 1181 GluIleSerCysValIySileGluAsnArgPhePheTyrGlnProSerProProThrCys 1200
Db 3541 GAGATCAAGTGTGGAAGATGAGAACAGAGTTCTTGTGGAGCCGACCCGACCAATATGC 3600
Qy 1201 IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
Db 3601 ATCCCTCCCTGCGGGGAGACCTTCAAGACCACTGTGAGATCACTCTCAACCAATATGC 3660
Qy 1221 ProGluProTyrProProGlyIyGluCysAspTyrIySValThrValSerProAspTyr 1240
Db 3661 CCAGAACCTTACCCGCGACGAGAGAGTGTGACTGGAAAGTACCGTCTCAACAGACTAC 3720
Qy 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
Db 3721 GTCAATCCCTCGGTATTTAACATCTTAACTGTGAGCTGTGCTATGACTCTCCCATATGC 3780
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeuIlegIySerPheTyrGlySerGlnLeuPro 1280
Db 3781 TACGACGAGAGGGAGCTCTCAAGCCCTCTCATAGAACTTTCTATGGCTCCAGACTCCA 3840
Qy 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GGGCGCATTTGAAAGACGACGACCAACAGCTCTCTCGCCCTCGGACGCAATGATGTGG 3900
Qy 1301 SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPheAspPro 1320
Db 3901 AGCAATGCTGGCTTGTCTCATTTGCTATACAGAAACCCCGGAGTCAAGTITTTATCTCT 3960
Qy 1321 GlySerIleIyAsnGlyThrArgValGlySerAspLeuIySleuGlySerSerValThr 1340
Db 3961 GGTTCATCAAGAACGAGCAGACGCGGTGGGGTCCGACCTGAAGCTGGGCTCTCCCTCAC 4020
Qy 1341 TyrTyrCysHisGlyValIyTyrGluValGluGlyThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTACTGCGACGGGGGCTTACAGAAAGTTGAGGACCTTCGACCTGAGCTCTCTGGGG 4080
Qy 1361 ProAspGlyIyAspProValTyrPheAsnProArgProValCysThrAlaProCysGlyVal 1380
Db 4081 CCGATGGGAGGCGGTGTGGAACAATCCCGGACAGTGTGACAGCCCTGTGGGGGA 4140
Qy 1381 GlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnTyrThrSer 1400
Db 4141 CAGTATGTGGTTCGACGAGAGTGTCTTGTCCCAACTACCCCGAGACTACACAGAT 4200
Qy 1401 GlyAlaIleCysLeuTyrPheValThrValProIyAspTyrValValPheGlyGlnPhe 1420
Db 4201 GGAAGATCTGCTGTGTATTTTGTACTGTCCCAAGAGACTATGTGTGTGGCCAGTTC 4260
Qy 1421 AlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGlnHis 1440
Db 4261 GCCTCTTTCACACGGCCCTCAACAGCGTGTGGAGTTCAAGACGCCACAGCCAGCAC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeuAlaThr 1460
Db 4321 TCGCGGCTCTCAAGCTCCCTCTCGGGCTCCATACAGAGAAATCACTGCTTGGCCAC 4380
Qy 1461 SerAsnGlnValLeuIleIySAspPheSerAlaIySgIyLeuAlaProAlaArgIyPheHis 1480
Db 4381 TCCAAATCAAGTTCTCATTTAAGTTACAGCCGCAAGGCTTGCACCGACGAGAGCTTCA 4440
Qy 1481 PheValTyrGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluPro 1500
Db 4441 TTTGTCTACAAAGCGTTCTCTGAACAGCGCCAGAGTGCAGCTGTGTCCGGAACCC 4500
Qy 1501 ArgTyrGlyIyAspArgLeuGlySerAspPheSerValGlyAlaIleValArgPheGluCys 1520
Db 4501 CCTATGGCAGAGAGCTGGGACAGTACTTCTCGGTGGGGCCATCTCGCTTCAATATGC 4560
Qy 1521 AsnSerGlyTyrAlaLeuGlnGlySerProGluIlegIyCysLeuProValProGlyAla 1540
Db 4561 AACTCCGCTATGCCCTCAAGGGGTCCGACAGATCGAGTCCCTCTGTGCTGGGGCC 4620
Qy 1541 LeuAlaGlnTyrAspValSerAlaProThrCysValValProCysGlyIyAsnLeuThr 1560
Db 4621 TTGGCCCAATGGAATGTCTACGCGCCACGVTGTGTGGCCGTGTGGAGGACCTTCA 4680
Qy 1561 GluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrIleAsnSerLeuAsn 1580
Db 4681 GAGCGCAGGGGACCATCTCTCCCTGTGCTTCCACAGCCGTACCTCAACAGCTTCAAC 4740
Qy 1581 CysValITyPlySileValValProGluGlyValagIyIlegIlnIlegIlnValSerPhe 1600
Db 4741 TGTGTGGAAGATCTGTGTCCCGAAGGCCCTGTGCATCCAGATCAAGTTGTCAGTTT 4800
Qy 1601 ValThrGluGlnAsnTyrAspSerLeuGluValPheAspGlyValaAspAsnThrValThr 1620
Db 4801 GTGACAGAGCAAGACTGGAGCTCCCTGAAATTTATAGGTGAGATTAACACTGTATACC 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGGGAGTTTCTCAGGAACAACGTGCTGCTTGTGAACAGACCTCCAAACAG 4920
Qy 1641 LeuTyrIleuHisPheTyrSerAspIleSerValSerAlaAlaGlyPheHisIleuGluTyr 1660
Db 4921 CTCTAATCTTCAATTTCTAAGATACAGATTCAGGATCTGCACCTGCTTCCACTTGGAGTAC 4980
Qy 1661 LysThrValGlyLeuSerSerCysProGluProAlaValaProSerAsnGlyValIySThr 1680
Db 4981 AAAACGGTGGGCTGAGACATTTCTCCGGAACCTGTGTGCCAGTAAAGGGGTGAAGACT 5040
Qy 1681 GlyGluArgTyrIleuValaAsnAspValaSerPheGlnCysGluProGlyTyrAlaLeu 1700
Db 5041 GCGCAGGCTTACTTGTGATATGATGTGTCTTTCAGAGTGTGAGCCGGGATATGCCCTTC 5100
Qy 1701 GlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTyrPheAsnTyrProPro 1720
Db 5101 CAGGGCCACGCCACATCTCTCAATGCCGGAACAGTGGGAGTGAAGAACTTCACTTCCCT 5160
Qy 1721 ProLeuCysIleAlaGlnCysGlyIyThrValGluGluMetGluGluValIleLeuSer 1740
Db 5161 CCACTGTATTCACAGATGTGGGGGACAGTGTGAGAGATGAGAGGGGTGATCTCGAAC 5220
Qy 1741 ProGlyPheProGlyIyAsnTyrProSerAsnMetAspCysSerTyrIySileAlaLeuPro 1760
Db 5221 CCGGCTTCCAGGCACTTACCCCAATGATGATGATGATGATGATGATGATGATGATGATGAT 5280

Qy	1761	ValGlyPheGlyValHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspThr	1780	Qy	2121	SerGluIleuSerGluIleuPheGluIlePheAspGlyProSerGlyGlnSerProLeu	2140
Db	5281	GTGGGCTTTGGAGCTCAACATCAAGTTCTGAACTTCTCCACGAGCCCAACAGACACTAC	5340	Db	6361	AGCAGAAAGCAATATGATGAGTTTGAGATTTTTCATGATGATCAGACAGAGTCTCTG	6420
Qy	1781	IleGluIleArgAsnGlyProTyrGluThrSerArgMetGlyArgPheSerGlySer	1800	Qy	2141	LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerAsnSer	2160
Db	5341	ATAGAATATCCGGAATGGCCCTATGAGACAGCCGCAATGATGGGAAGTTCAAGTGAAGC	5400	Db	6421	CTGAAGCCCTCAAGTGGGAATTAATCAAGCTCCCTGATTTGTCACCACTCAAGCACTCT	6480
Qy	1801	GluLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTyrPheHisSerAsp	1820	Qy	2161	ValTyrLeuArgTyrPheSerSerAspHisAlaTyrAsnArgIleGlyIlePheValIleArgTyr	2180
Db	5401	GAGCTTCAGACTCCCTCTCTCCACGTCACGAGACACCGTGATTTTCCACAGCGCAC	5460	Db	6481	GTTGATCTGGTGGTGTATGTATGACGCTTACATGGAAGGGCTTCAAGATCCGCTAT	6540
Qy	1821	HisSerGlnAsnArgProGlyPheLeuLeuGluTyrGlnAlaTyrGluLeuGluGlnCys	1840	Qy	2181	SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlnThr	2200
Db	5461	CACCTCCCAATCGGACGAGATCAAGCTGAGATCAAGGCTTGAATCTTCAAGAGAGTC	5520	Db	6541	TCAACCTTATCTGACCTGACCTGACCGGCTTCACTTCACTGAGCTTCACTTCAAGCCAAACC	6600
Qy	1841	ProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGln	1860	Qy	2201	SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly	2220
Db	5521	CCAGACCCAGAGCCCTTTGCCAATGGCATTTGAGGGGAGCTGGCTACAACTGGAGCAA	5580	Db	6601	AGACCCAGCCGGGGCTCCATCACTTGGGCTGCAAGCCGGCTTACCGGCTGGTGGGA	6660
Qy	1861	SerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProValLeuThrCys	1880	Qy	2221	HisSerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrPheSerGluAlaIle	2240
Db	5581	TCAGTGAACCTTCGAGTGCCTCCGGGGTATCAATTGACTGGCCACCTGTCTCAGCTGT	5640	Db	6661	CACAGCATGGCCATCTGTATCCCGCACCCAGGGCTTACCACTGTGGAGCGAAGCATC	6720
Qy	1881	GlnHisGlyThrAsnArgAsnTyrAspHisProLeuProLysCysGluValProCysGly	1900	Qy	2241	ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPhe	2260
Db	5641	CAACATGGCACCAACCGGAACTGGAGAACCCCTGCCAAGTGAAGTCCCTTGTGGC	5700	Db	6721	CTCTCTGTCAAGCTCTTCTCTGTGGGCTTCTGAGGCCCCCAAGAAATGATGTGTT	6780
Qy	1901	GlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSer	1920	Qy	2261	GlyLysGluTyrThrValGlyThrValAlaValTyrSerCysSerGluGlyTyrHisLeu	2280
Db	5701	GGGAACATCACTTCTTCCACGAGCATGTGTACTCCCGGGGTTCCCTAGCCGCTACTCC	5760	Db	6781	GGCAAGAGATACACAGTGGGAACCAAGCCATGATACAGCTGAGTGAAGGCTACCACTC	6840
Qy	1921	SerSerGlnAspCysValTyrLeuIleThrValProIleGlyHisGlyValArgLeuAsn	1940	Qy	2281	GlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuTyrPheSerAsnArgAsn	2300
Db	5761	AGCTCCACGAGACTGTGTGCTGCTGATCACCTGGCCCATTTGGCCATGGGTCGCGCTCAAC	5820	Db	6841	CAGGACAGCCCTGAGGCGCATGCAAGTGTCTGAGACAGAGCTTATGACCAACCGCAAT	6900
Qy	1941	LeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAsnGlyProGln	1960	Qy	2301	ValProProGlnCysVal-----ProValThrCysPro	2311
Db	5821	CTTACGCTGTGACAGACAGAGCCCTGTGAGATTTCAACACATCTGGAGTGGGCCACAG	5880	Db	6901	GTCCACCAAGTGTGTCTGTGAGTCTCTGGGCAATGAGCGGGGTGTGATCTTGCT	6960
Qy	1961	GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLysThrValGlnSer	1980	Qy	2312	AspValSerSerIleSerValGluHisGlyArgTyrArgLeuIlePheGluThrGlnTyr	2331
Db	5881	CAACAGAGACCAAGGCTCGGCTCTTCAACCGAGCATGGCCAAAGAAACGTCCACAGT	5940	Db	6961	GATGTCAGTACATCAAGCTGAGAGCATGGAGCATGGCGATGAGGCTTATCTTGAACACATAT	7020
Qy	1981	SerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAlaThrGlyIlePheAla	2000	Qy	2332	GlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGlyGlnArg	2351
Db	5941	TCATCCCAACGAGTCTGCTCAAGTTCACCGTATGAGCCACAGGGGGGATCTTGCC	6000	Db	7021	CAGTTCAGGCGCCAGCTGATGTATCTGTGACCTGGCTGATCACTATCTGCGCAAGG	7080
Qy	2001	IleAlaPheSerAlaTyrProLeuThrLysCysProProThrIleLeuProAsnAla	2020	Qy	2352	ValIleArgCysGlnAlaAsnGlyLysTyrPheSerLeuGlyAspSerThrProThrCysArg	2371
Db	6001	ATACCTTCTCTCGCTTATCCATCAACCAATGCTCTCTCCACCATCTCTCCCAAGCC	6060	Db	7081	GTATCCGCTGTCAAGCCAAATGGCAATGGAGCTCGGGGACTTACGCCACCTTGCGA	7140
Qy	2021	GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeu	2040	Qy	2372	IleIleSerCysGlyGluLeuProIleProProAsnGlyHisArgIleGlyThrLeuSer	2391
Db	6061	GAACTCGTCAACAGATGAAATTAATTAAGTACATCTGACCTGACCTACAGTATGCTCTC	6120	Db	7141	ATCATCTCTGTGAGAGCTCCCGATTTCCCCCAATGGCCACCGCATGGAAACACTGTCT	7200
Qy	2041	ProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGln	2060	Qy	2392	ValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrLeuValGlySerArg	2411
Db	6121	CTGGCTTATCTTAACTGGGGAATGAATTTCTGACCTGCAAACTTGAACCTTACTCTCAG	6180	Db	7201	GTCACGGGGCAACAGCATCTTCTCTGCAATTCGCGAATACACACTGTGGGCTCCAGG	7260
Qy	2061	PheGluGlyProProProIleCysGluValHisCysProThrAsnGluLeuLeuThrAsp	2080	Qy	2412	ValArgGluCysMetValaAsnGlyLeuTyrPheSerGlySerGluValArgCysLeu	2429
Db	6181	TTTAAAGGACCAACCCCGAATATGAAATGACATGTCCAACTGAAGAGCTTCTGACAGAC	6240	Db	7261	GTCGTGATGATGATGGCAATGGGCTCTGAGTGGCTGTGAAGTCCGCTCTTGCCACT	7320
Qy	2081	SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys	2100	Qy	2429	-----	2429
Db	6241	TTCCAGAGCGGTGATCTGAGCCAGACACTACCTTGAAAGCTATCCCAAGTTCAGACTTC	6300	Db	7321	CAGACCAAGCTCACTCCATTTTCTATTAAGCTCTCTGTGATGATCACTCTTCCCATCC	7380
Qy	2101	SerTyrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu	2120	Qy	2430	-----AlaGlyHisCysGlyThrProGluProIleValaAsnGlyHisIleAsnGly	2446
Db	6301	TCTTGCGTGTGAGAGTGAAGCCGACACTATTAACATCTCCCTCAAGAGTGAATCTTCTC	6360	Db	7381	CTCACAAGAACTGTGACACTGTGGAGACTCTCGAAGCCCATTTGTCAACGGACACATCAAT	7440
Qy	2121	TCCTTGCGTGTGAGAGTGAAGCCGACACTATTAACATCTCCCTCAAGAGTGAATCTTCTC	6360	Qy	2447	GluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIle	2466

PR	07-JUN-2002;	2002US-0386531P.
PR	07-JUN-2002;	2002US-0386942P.
PR	07-JUN-2002;	2002US-0386971P.
PR	07-JUN-2002;	2002US-0387262P.
PR	08-JUN-2002;	2002US-0296660P.
PR	10-JUN-2002;	2002US-0387400P.
PR	10-JUN-2002;	2002US-0387535P.
PR	11-JUN-2002;	2002US-0387610P.
PR	11-JUN-2002;	2002US-0387625P.
PR	11-JUN-2002;	2002US-0387634P.
PR	11-JUN-2002;	2002US-0387668P.
PR	11-JUN-2002;	2002US-0387696P.
PR	11-JUN-2002;	2002US-0387702P.
PR	11-JUN-2002;	2002US-0387836P.
PR	11-JUN-2002;	2002US-0387859P.
PR	12-JUN-2002;	2002US-0387933P.
PR	12-JUN-2002;	2002US-0387934P.
PR	12-JUN-2002;	2002US-0387960P.
PR	12-JUN-2002;	2002US-0388022P.
PR	12-JUN-2002;	2002US-0388066P.
PR	13-JUN-2002;	2002US-0389123P.
PR	14-JUN-2002;	2002US-0389118P.
PR	14-JUN-2002;	2002US-0389120P.
PR	14-JUN-2002;	2002US-0389144P.
PR	14-JUN-2002;	2002US-0389146P.
PR	17-JUN-2002;	2002US-0389729P.
PR	17-JUN-2002;	2002US-0389742P.
PR	18-JUN-2002;	2002US-0389884P.
PR	19-JUN-2002;	2002US-0390006P.
PR	19-JUN-2002;	2002US-0390209P.
PR	21-JUN-2002;	2002US-0390763P.
PR	17-JUL-2002;	2002US-0396706P.
PR	06-AUG-2002;	2002US-0401628P.
PR	09-AUG-2002;	2002US-0402156P.
PR	09-AUG-2002;	2002US-0402256P.
PR	12-AUG-2002;	2002US-0402389P.
PR	12-AUG-2002;	2002US-0402786P.
PR	12-AUG-2002;	2002US-0402816P.
PR	12-AUG-2002;	2002US-0402821P.
PR	13-AUG-2002;	2002US-0402832P.
PR	13-AUG-2002;	2002US-0403448P.
PR	13-AUG-2002;	2002US-0403459P.
PR	13-AUG-2002;	2002US-0403531P.
PR	13-AUG-2002;	2002US-0403532P.
PR	13-AUG-2002;	2002US-0403563P.
PR	13-AUG-2002;	2002US-0406317P.
PR	15-AUG-2002;	2002US-0406317P.
PR	26-AUG-2002;	2002US-0406182P.
PR	26-AUG-2002;	2002US-0406356P.
PR	27-AUG-2002;	2002US-0406240P.
PR	12-SEP-2002;	2002US-0410084P.
PR	20-SEP-2002;	2002US-0412528P.
PR	23-SEP-2002;	2002US-0412731P.
PR	30-SEP-2002;	2002US-0414801P.
PR	30-SEP-2002;	2002US-0414839P.
PR	30-SEP-2002;	2002US-0414954P.
PR	09-OCT-2002;	2002US-0417186P.
PR	09-OCT-2002;	2002US-0417406P.
PR	23-OCT-2002;	2002US-0420639P.
PR	28-OCT-2002;	2002US-0421156P.
PR	31-OCT-2002;	2002US-0422690P.
PR	01-NOV-2002;	2002US-0423110P.
PR	05-NOV-2002;	2002US-0423379P.
PR	05-NOV-2002;	2002US-0423798P.
PR	12-NOV-2002;	2002US-0425453P.
XX	(CURA-) CURAGEN CORP.	
PA		
XX		
EI	Alsbrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SU;	
EI	Catterton E, Chapoval A, Cabriere-Bokor JR, Edinger SR, Ellerman K;	
EI	Brenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;	
EI	Gusev VY, Hermann II, Ti W, Koch J,	

QY 521 ThrIleLeuSerProGlyTyrProGlyPheTyrIleAspAlaLeuSerCysAlaTyrVal 540
 DB 1561 ACCATCTCTCTCCGGGCTGAGCTTCAAGAGATGCTTGAGCTGCTGGAGTG 1620
 QY 541 IleguaIagInProGlyTyrProIleValIleThrPheAspArgPheIleThrGluVal 560
 DB 1621 ATTGAGGCCAGGCGAGCTACCCCATCAAAATCACTTCGACAGATTCAAAACCGAGAGTC 1680
 QY 561 AsnTyrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaProLeuIleGlyVal 580
 DB 1661 AACTATGACACCTCGAAGTACGCGATGGCGGACTTACAGCGCCCTTGATCGGGGTT 1740
 QY 581 TyrIleGlyThrGluValProGluPheLeuIleSerThrSerAsnTyrLeuTyrLeuLeu 600
 DB 1741 TACGACGGGACCCAGGTTCCCAAGTCTCTCATCAGACACGACCACTTCTTCTCTC 1800
 QY 601 PheSerThrAspIleSerHisSerAspIleGlyPheGluLeuArgTyrGluThrIleThr 620
 DB 1801 TTCTCTACCGACAAGAGTCACTCGGACATCGGCTTCCAGCTCCGCTATGAGACTATACA 1860
 QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyValAsnGlyValAsn 640
 DB 1861 CTGACGTAGACCACTGCTGTGATCCAGAAATCCAGTAATGACAGCGTCAAGGGAAAT 1920
 QY 641 AspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeuSerAsp 660
 DB 1921 GACTTTCATCGTGGGCGGCTGTGACCTTGACGTGACCTCGGCTACACATTAAAGTAC 1980
 QY 661 GlyGluProLeuGluCysGluProAsnPheGlnTyrSerArgAlaLeuProSerCysGlu 680
 DB 1981 GGGAGCCTCTGGAGGTGAGCCCACTTCAGTGAAGCGGGCCCTGACAGTTTGAA 2040
 QY 681 AlaLeuCysGlyGlyPheIleGlnIleGlySerSerGlyThrIleLeuSerProGlyPhePro 700
 DB 2041 GCTCTGTGTGGTCTCATTTCAAGGCTCCAGTGGAGCAATCTTGCGCAAGGATTCCT 2100
 QY 701 AspPheTyrProAsnAsnLeuAsnCysThrTyrIleIleGluThrSerHisGlyValGly 720
 DB 2101 GACTTTCATCCCAACAACCTGAACTGCACTGGAATTCGAAACATCTCATGCGAAGGGT 2160
 QY 721 ValPhePheThrPheHisThrPheHisAlaLeuGluSerGlyHisAspTyrLeuLeuIleThr 740
 DB 2161 GTGTTCTTCACTTTCCACACCTTCCACCTGAAAGTGGCATACACTCCCTCATCACT 2220
 QY 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
 DB 2221 GAGAACGGGAGCTTCAACCCAGCCCTGAGGCACTAACTGATCTCGGCTGCCAGCTCCC 2280
 QY 761 IleSerIleGlyLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
 DB 2281 ATCAGCGCTGGGCTCTATGCGAACTTCACCTGACCAGGTCCGCTTCACTCTGATTTCTCC 2340
 QY 781 MetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCysGluGlu 800
 DB 2341 ATGTCAATATGAAGATTCAACATCACTTCTGAGATGACACTTGAGGCCCTGTGAGGAG 2400
 QY 801 ProGluValProAlaTyrSerIleArgGlyGlyLeuGlnPheGlyValGlyAspThrLeu 820
 DB 2401 CCCGAGGTCCACACTCAAGCATCCGGAAGGGCTTCAGTTGGCGTGGCGACACCTTG 2460
 QY 821 ThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCysLeuGly 840
 DB 2461 ACCTTCTCTGCTTCCCGGGATCCGCTCGAGGGGACCGCCGACATCAGTCCGGGGG 2520
 QY 841 GlyArgArgArgLeuTyrSerSerProLeuProArgCysValAlaGluCysGlyValSer 860
 DB 2521 GGCAGACGGGCGCTGAGAGCTGCTGCCAAGGTGTGTGCTGAGTGGAGATTCA 2580
 QY 861 ValThrGlyThrGlnGlyThrIleLeuSerProAsnPheProValAsnTyrAsnAsn 880
 DB 2581 GTCAAGGACATCAAGGTACTTGTGTCTCCCACTTCTGTGAACATAATACAT 2640

QY 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGlnLeuValArg 900
 DB 2641 CATGAATGATCACTCACTCCATCCAGACCCAGGAGGAGGAATTCACTGAAGCCAG 2700
 QY 901 AlaPheGluLeuSerGluGlyAspValLeuValValTyrAspGlyAsnAsnAsnAla 920
 DB 2701 GCATTGAACTCTCCAGAGAGATGTCTCAAGTTTATGATGGCAACAACATCCGCC 2760
 QY 921 ArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsnSerThrSer 940
 DB 2761 CGTTTCTGGAGATTTTTATGCCATTCTGAGATGATGGGGGTGACTTTGAACGACATCC 2820
 QY 941 SerSerLeuTyrLeuAspPheIleThrAspAlaGluAsnThrSerIleGlyPheGluLeu 960
 DB 2821 AGCAGTGTGCTGATTTCACTCATGATGCTGAACAACCCAGCAGAGGCTTTGAACAG 2880
 QY 961 HisPheSerSerPheGluLeuIleIleCysGlyAspProGlyThrProIlePheGlyTyr 980
 DB 2881 CACTTTCCAGCTTTGAACTCATCAATGTGAGAACCCAGAAACCCCAAGTTGGCTAC 2940
 QY 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCysAspProGly 1000
 DB 2941 AAGGTTCAATGATGAAGTCAATTTTGACAGGAGCTCCGTGCTTCAAGCTGTGACCTG 3000
 QY 1001 TyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyValArgThrTyrAsp 1020
 DB 3001 TACAGCTGCGGGAGTGAAGAGCTGTGTGTGATGAGAGAGCGGACCTGGAGAC 3060
 QY 1021 ArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyValSerGly 1040
 DB 3061 CGGCTCTGCCACCTGTGTGCGCGAGTGGAGGACAGTGAAGAGAGAGTGTGCGGG 3120
 QY 1041 GlnValIleLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleTyrThr 1060
 DB 3121 CAGGTGTGTACCCCGGATATCCAGCTCCATATGAACAACATCTCACTGATCGGACC 3180
 QY 1061 IleguaIagIleValCysThrIleGlyLeuHisAspLeuValPheAspThrGluGlu 1080
 DB 3181 ATCGAAGAGAGGCGGCTGACCATTTGGCTTACACTTCTGTGTGTGAGACAGAGAG 3240
 QY 1081 ValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyValLeuLeuValGlu 1100
 DB 3241 GTTCACGAGTGTGCGCATCTGGAGATGGGCTGTGAGAGACGGGTTCTGTGAAAGAG 3300
 QY 1101 LeuSerGlyProAlaLeuProIleAspLeuHisSerThrPheAsnSerValIleGln 1120
 DB 3301 CTGAGTGGCCGGGCTGCGCCAGAGACCTGATAGCACTTCAACCTGCTGCTCGAG 3360
 QY 1121 PheSerThrAspPhePheThrSerIleGlnIlePheAlaIleGlnPheSerValSerThr 1140
 DB 3361 TTCAGCACTGACTTCTTACAGCAAGACAGAGGCTTTGCCATTTCAATTTTCAAGTGT 3420
 QY 1141 AlaThrSerCysAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr 1160
 DB 3421 GCAAGTCTCCGATGACCTGCGATCCCGCAGATGGAGTGGAGTGTGACAGATTGG 3480
 QY 1161 GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrValAlaLeuGlnGlySerAla 1180
 DB 3481 GAACCCGGCACTCCACAGTGTTCATGATGACCTTGCTACCGGCTGCGAGGAGAGTGA 3540
 QY 1181 GluIleSerCysValIleIleguaAsnArgPhePheTyrGlnProSerProProThrCys 1200
 DB 3541 GAGATCAGCTGTGTGAAGATCGAAGACAGGTTCTTGTGGACGCCAGCCGCAACATGC 3600
 QY 1201 IleAlaProCysGlyValAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
 DB 3601 ATGCTCTCCGCGGGAGAGCTGACAGACCAATCTGAGATCACTCTTCAACCAATTAC 3660
 QY 1221 ProGluProTyrProProGlyValGluCysAspTyrIleValThrValSerProAspTyr 1240
 DB 3661 CAGAACTCAACCGCCAGGCAAGAGTGTGATGAAAGTGAACCTTCAACCAAGACTAC 3720
 QY 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrTyrAspPheLeuHisIle 1260

```
Db 3721 GTCAATCGCCCTGGATTTTAACATCTTTAACTGGAGCCTGGGTATGACTTCTCCATATC 3780
Qy 1261 TTTATPAGLYATGAPSerLeuSerProLeuIleGlySerPheTYRGLYSerGlnLeuPro 1280
Db 3781 TACAGCGGACGGGACTCTCTCAGCCCTCTCAAGAACCTTCTATGGCTCTCCAGCTCCCA 3840
Qy 1281 GLYATGILeGlySerSerSerSerPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GGCGGCATGTGAAGACAGACAGACAGCTTCTCTCGCCTTCCGAGGAGTGCATCTGTG 3900
Qy 1301 SerAsnAlaGlyPheValIleAspTYRThrGlnAsnProArgGlySerCysPheAspPro 1320
Db 3901 AGCAATGCTGGCTTGTCATTGACTATACAGAAAACCGCGGAGTCAATGTTTGAATCT 3960
Qy 1321 GlySerIleLYPAsnGlyThrArgValIleGlySerAspLeuLYPLeuGlySerSerValThr 1340
Db 3961 GGTTCCATCAAGAACGGGACACGGGTGGGTCCGACCTGAAGCTGGGCTCTCCGTCAAC 4020
Qy 1341 TYTTYRCSBHsgLYGLYTYRGLYValIleGlyThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTATCTCCACGGGGCTACGAGATGAGGACCTGACCTGAGCTGCACTCTGAGG 4080
Qy 1361 ProAspGlyLYPProValITRPAsnAsnProArgProValCysThrAlaProCysGlyGly 1380
Db 4081 CCTGATGGGAAGCCCGTGTGGAACAATCCCGGCGCAGCTGCACAGCCCTGTGGGGA 4140
Qy 1381 GlnTYRValIleGlySerAspGlyValIleLeuSerProAsnTYRProGlnAsnTYRThrSer 1400
Db 4141 CAGTATGGGTGGTCCGACGGAGGTGTTGTCTCCCACTACCCCAACATACACCACTAC 4200
Qy 1401 GLYGlnIleCysLeuTYRPhenValThrValProLYSAspTYRValIlePheGlyGlnPhe 1420
Db 4201 GGACGATCTGCTGTATTTGTTATCTGTGCCAAAGACTATGTGTGTTGGCCAGTTC 4260
Qy 1421 AlaPhePheIleThrAlaLeuAsnAspValIleGlyValIleAspGlyIleSerGlnIle 1440
Db 4261 GCCTTCTTTCAACGGGCCCTCAAGACGTGTGGAGGTTCACAGAGGCCACAGCCAGAC 4320
Qy 1441 SerArgLeuLeuSerSerLeuSerGlySerIleThrGlyGlySerLeuProLeuAlaThr 1460
Db 4321 TCGCGGCTCTCCAGCTCCCTCTCGGGCTCCCATACAGAGAAATCACTCCCTGGCCAC 4380
Qy 1461 SerAsnGlnValIleuIleLYSPheSerAlaLYSGLYLeuAlaProAlaArgGlyPheIle 1480
Db 4381 TCCATCAAGATTCTCATTAAGTTCAAGCCCAAGGCCCTGCACACGACGAGGCTTCCAC 4440
Qy 1481 PheValTYRGLNAlaValProArgThrSerAlaThrGlnCysSerSerValProGlnPro 1500
Db 4441 TTTGCTACCAAGCGGTTCTCTCGAACCAAGCCGACAGCTGCTGTCCGGAAACC 4500
Qy 1501 ArgTYRGLYLYPArgLeuGlySerAspPheSerValIleGlyAlaIleValArgPheGlnCys 1520
Db 4501 CGCTATGGCAAGAGGCTGGGAGTGAATCTCTCGGTGGGGCCATCTGTCCGATTCGAAATGC 4560
Qy 1521 AsnSerGlyTYRAlaLeuGlnGlySerProGlnIleGlyCysLeuProValProGlyAla 1540
Db 4561 AACTCCGGCTATGCCCTGCAGAGGGTGCAGAGATCGAGTCCCTCCGTCCCTGGGGCC 4620
Qy 1541 LeuAlaGlnTRPAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeuThr 1560
Db 4621 TTGGCCCAATGGAATGTCTACGCGCCACGCTGTGTGGTCCGTGTGAGGACACTCA 4680
Qy 1561 GlnArgArgGlyTYRThrIleLeuSerProGlyPheProGlnProTYRLeuAsnSerLeuAsn 1580
Db 4681 GAGGCGAGGGGACACCATCTGTCTCCCTGCTTCCAGAGCCGTAACCAAGCCTCAAC 4740
Qy 1581 CysValITRPLYSerIleValValProGlnGlyValIleGlnIleGlnValIleValSerPhe 1600
Db 4741 TGTGTGTGAAGATGTGTGTCCCGAAGCCGTGCATTCAGATCCAAAGTTGTCAAGTTT 4800
Qy 1601 ValThrGlnGlnAsnITRPApSerLeuGlnValPheAspGlyAlaAspAsnThrValThr 1620
Db 4801 GTGACAGACAGAACTGGACTCGCTGGAATATTGATGTGCAGATTAACACTGTAAAC 4860
Qy 1621 MetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsnGln 1640
Db 4861 ATGCTGGAGAGTTTCTCAGAAACAACCGTGTCCCTTCTGAAACACACACTTCAACACG 4920
Qy 1641 LeuTYRLeuIlePheTYRSerAspIleSerValSerAlaIleGlyPheIleLeuGlyTYR 1660
Db 4921 CTCACTTCACTTTCTACTCAGATATCAGCTATCTGACACTGGCTTCCACTTGGATAC 4980
Qy 1661 LYSPheValIleLeuSerSerCysProGlnProAlaValProSerAsnGlyValIleThr 1680
Db 4981 AAAACGGGTGGCCTGAGACAGTGTCCGAAACCTGCTGTGCCACAGTAACGGGGTGAAGCT 5040
Qy 1681 GLYGlnArgTYRLeuValIleAsnAspValIleSerPheGlnCysGlyPProGlyTYRAlaLeu 1700
Db 5041 GGCGAGGCTACTTGGTGAATGAATGTGTGTCTTCCAGTGTGCGGGAGTATGCTCTC 5100
Qy 1701 GlnGlyIleAlaIleSILESerCysMetProGlyTYRValIleArgArgTRPAsnTYRProPro 1720
Db 5101 CAGGGCCACGCCCATCTCTCTGCATGCCGGAACAGTGGGCGCATGGAATCACTCTCT 5160
Qy 1721 ProLeuCysIleAlaGlnCysGlyGlyTYRThrValIleGlnLeuGlyValIleLeuSer 1740
Db 5161 CCACTTGATATTGCAGAGTGTGGGGAACAGTGGAGAGATGGAGGGGTGATCTGAGC 5220
Qy 1741 ProGlyPheProGlyValAsnTYRProSerAsnMetAspCysSerTrpLYSLeuAlaLeuPro 1760
Db 5221 CCCGGCTTCCAGGCAACTACCCAGTAAATGAGTACCTCCGGAATAAATAGCACTGCC 5280
Qy 1761 ValIlePheGlyAlaIleSILEGlnPheLeuAsnPheSerThrGlnProAsnIleAspTYR 1780
Db 5281 GTGGGCTTTGGAGTCAATCACTCACTCTTGAATCTTCCACCGAGGCCAACAGACTAC 5340
Qy 1781 IleGlnIleArgAsnGlyProTYRGLYThrSerArgMetGlyArgPheSerGlySer 1800
Db 5341 ATAGAATCCGGAAATGGCCCTTATGAGACAGCCGATGAGGAGATTCAATGAGGAAGC 5400
Qy 1801 GlnLeuProSerSerLeuLeuSerThrSerIleGlyThrValTYRPhenIleSerAsp 1820
Db 5401 GAGCTTCAAGCTCCCTCTCTCCACGTCCACAGACCAACCGGTATTTTCCACAGGAC 5460
Qy 1821 HisSerGlnAsnArgProGlyPheLYPLeuGlyTYRGLNAlaTYRGLYLeuGlnGlyCys 1840
Db 5461 CACTTCCAGATCGCCAGAGATTCAAGCTGAGATAGAGCTTATGAATCTTCAAGATGCC 5520
Qy 1841 ProAspProGlnProPheAlaAsnGlyIleValArgGlyAlaGlyTYRAsnValGlyGln 1860
Db 5521 CCAGACCCAGAGCCCTTTGGCAATGGCATTTGTGAGGGAGCTGGCTTCAACGTGGGCAA 5580
Qy 1861 SerValIThrPheGlyCysLeuProGlyTYRGLNLeuThrGlyIleProValIleuThrCys 1880
Db 5581 TCAGTGACTTTCAGTGTCTCCCGGGGTATCAATGACTGGCCACCTGTCTTCCACGTGT 5640
Qy 1881 GlnIleGlyIleThrAsnArgAsnITRPApIlePheLeuProLYSCysGlyValProCysGly 1900
Db 5641 CAACATGACACCAACCGAACTGGAGACACCCCTGCCAAGTGTGAATCCCTGTGGCC 5700
Qy 1901 GLYAsnIleThrSerSerAsnGlyTYRValTYRSerProGlyPheProSerProTYSer 1920
Db 5701 GGGAACATCACTTCTTCCAAAGGACAGTGTATCTCCCGGGGTTCCCTTACCCGTAATCC 5760
Qy 1921 SerSerGlnAspCysValITRPLeuIleThrValProIleGlyIleGlyValIleArgLeuAsn 1940
Db 5761 AGCTCCAGAGACTGTGCTGTGATGATCAACGTGCCATTTGCCATGGCCGTCCCTCAAC 5820
Qy 1941 LeuSerLeuGlnGlnThrGlnProSerGlyAspPheIleThrIleTRPApGlyProGln 1960
Db 5821 CTGAGCTGCTGCAAGACAGAGCCCTCTGGAGATTTATCAACATCTGGAGTGGCCACAG 5880
Qy 1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLYSlyThrValGlnSer 1980
Db 5881 CAAACAGACACAGGCTCGGCGTCTTCACCCGGAGCATGGCCAAAGAAAACAGTGCAGAGT 5940
```

```
QY 1981 SerSerAsnGlnValLeuLeuPheHisArgAspAlaAlaThrGlyGlyIlePheAla 2000
DB 5941 TCATCCAAACAGGCTCTGCTCAAGTTCCACCGTGAATGACGCCACAGGGGGGATCTTCC 6000
QY 2001 IleAlaPheSerAlaTyrProLeuThrIleCysProProProThrIleLeuProAsnAla 2020
DB 6001 ATAGCTTTCTCCGCTTATCCTCACTCAACAAATGCCCTCCCTCCACCATCTCCCAAGCC 6060
QY 2021 GlnValValThrGlnAsnGlnGluPheAsnIleGlyAspIleValArgTyrArgCysLeu 2040
DB 6061 GAAGTCGTACAGAGAAATGAGAAATTCATATAGGTGACATCGATGACGTACAGATGCTCTC 6120
QY 2041 ProGlyPheThrLeuValGlyAsnGlnIleLeuThrCysIleGlyThrTyrLeuGln 2060
DB 6121 CTGCTTATCTTCTAGTGGGGAATGAAATTCGACCTCCAACTTGGAACCTTCACTGAG 6180
QY 2061 PheGlnGlyProProProIleCysGlnValHisCysProThrAsnGlnLeuLeuThrAsp 2080
DB 6181 TTGGAAGACACACCCCGATATGTGAAAGTGAAGTGTCCACAAATGAGCTTCTGACAGAC 6240
QY 2081 SerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCys 2100
DB 6241 TCACAGGCGGTGATCTCTAGCCAGAGCTAACCTTGAAAGCTAACCCAGTTCCAGACCTGC 6300
QY 2101 SerThrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeu 2120
DB 6301 TCTTGCGCTGAGAGTGGAGCCGACCTAATCACTCCCTCCACAGTGGAGTACTTCTC 6360
QY 2121 SerGlnGlyGlnTyrAspGluPheGlnIlePheAsnGlyProSerGlyGlnSerProLeu 2140
DB 6361 AGCGAAGACCAATATGATGATGATTTGAGATTTATGATGATGATGATGATGATGATGATG 6420
QY 2141 LeuValAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSer 2160
DB 6421 CTGAAAGCGCTCACTGGGGAATTACTCAGCTCCCTGATTTGTCACACAGCTCAAGCACTC 6480
QY 2161 ValTyrLeuArgTyrPheSerSerAspHisAlaTyrAsnArgIleGlyPheIleArgTyr 2180
DB 6481 GTGTAACCTGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6540
QY 2181 SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr 2200
DB 6541 TCAGCCCTTACTGCACTGCTCCAGGCTCCAGGCTCCAGTGGCTTCACTTCACTGAGCCAG 6600
QY 2201 SerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly 2220
DB 6601 AGCACCCAGCCCGGGGCTCCATCTCACTTGGCTGCAACGCGGCTAACCGCTGAGGGA 6660
QY 2221 HisSerMetAlaIleCysThrArgHisArgProGlnGlyTyrHisLeuThrPheSerGln 2240
DB 6661 CACAGCATGGCCCATCTTACCCGGACCCCGAGGCTACACCATGTCGAGGCGAAGCCATC 6720
QY 2241 ProLeuCysGlnAlaLeuSerCysGlyLeuProGlnAlaProIleAsnGlyMetValPhe 2260
DB 6721 CCTCTGCTCAAGCTCTTCTCTGAGGCTCTCGAGGCCCCCAAGAAATGATGATGATGAT 6780
QY 2261 GlnIleGlnIleTyrThrValGlyThrIleValAlaTyrSerCysSerGlnGlyTyrHisLeu 2280
DB 6781 GGGAGGAGATGACAGAGGGAGAACCAAGGCGCATGTACAGCTGACGTGAAAGGCTACCACT 6840
QY 2281 GlnAlaGlyAlaGlnAlaThrAlaGlnCysLeuAspThrGlyLeuThrPheSerAsnArg 2300
DB 6841 CAGGACGGGCGCTGAGGCGCATGTCAGAGTGTGACACAGGCTTATGAGCAACCGCAT 6900
QY 2301 ValProProGlnCysVal-----ProValThrCysPro 2311
DB 6901 GTCCCAACCAAGTGTCTCGTGAAGTCTCGGCAATGAGAGCGGAGTGTGATGATGATGAT 6960
QY 2312 AspValSerSerIleSerValGlnHisGlyArgTyrPheGlnIlePheGlnThrGlnTyr 2331
DB 6961 GATGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7020

QY 2332 GlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyrTyrThrGlnArg 2351
DB 7021 CAGTTCCAGGCGCCAGCTGATGCTCATCTGATGACCTGGCTACTATATCTGGCAAAAG 7080
QY 2352 ValIleArgCysGlnAlaAsnGlyIleTyrPheSerLeuGlyAspSerThrProThrCysArg 2371
DB 7081 GTCATCGCTGTGAGGCGCAATGAGCAATGAGCTCTGGGAGCTTACGCGCCACCTGCCA 7140
QY 2372 IleIleSerCysGlyGlnLeuProIleProProAsnGlyHisArgIleGlyThrLeuSer 2391
DB 7141 ATCATCTCTGTGAGAGCTCCCGATTCGCCCAATGGCCACCGCATCGGAACCTGTCT 7200
QY 2392 ValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThrLeuValGlySerArg 2411
DB 7201 GTCTACGAGGCAACACACCATCTTCTCCGCAATTCGAGATACACACTGAGTGGGCTCCAG 7260
QY 2412 ValArgGlnCysMetAlaAsnGlyLeuThrPheSerGlySerGlyValArgCysLeu----- 2429
DB 7261 GTGCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7320
QY 2429 ----- 2429
DB 7321 CAGACCAAGCTCACCTCATTTTCTAATAAGCTCTCTTGATGATGATGATGATGATGAT 7380
QY 2430 -----AlaGlyHisCysGlyThrProGluProIleValAlaAsnGlyHisIleAsnGly 2446
DB 7381 CTCACCAAGCTGAGACCTGTGGACCTCTGAGCCCATTTGTCACAGACATCAATAGG 7440
QY 2447 GlnAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIle 2466
DB 7441 GAGAACTACAGCTACCGGGGAGTGTGTGATACCAATGCAATGCGTGTCCGCTGATC 7500
QY 2467 GlyMetSerValArgIleCysGlnGlnAspHisHisTyrPheSerGlyIleThrPheCys 2486
DB 7501 GGCATGTCTGTGCGATCTGCACAGAGATATCATCTGTGTGGGCAACCCCTTCTGT 7560
QY 2487 Val----- 2487
DB 7561 GTGCATGTTAAGACAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGATGATGATGAT 7620
QY 2488 -----ProIleThrCysGlyHisAspProGlyAsnProValAsnGlyLeu 2501
DB 7621 GAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7680
QY 2502 ThrGlnGlyAsnGlnPheAsnLeuAsnAspValValIlePheValCysAsnProGlyTyr 2521
DB 7681 ACTCAGGCTAACCACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 7740
QY 2522 MetAlaGlnIleValAlaArgSerGlnCysLeuAlaSerGlyGlnTyrPheSerAspMetLeu 2541
DB 7741 ATGCTGTGAGGGGCTGCTAGGTCCCAATGCTGCGCAGCGGGCAATGAGTGAATCTG 7800
QY 2542 ProThrCysArgIleIleLeuAsnCysThrAspProGlyHisGlnGlnAsnSerValArgGln 2561
DB 7801 CCCACCTGCAAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 7860
QY 2562 ValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArgCysAsn 2581
DB 7861 GTCCACGCGACGCGCGCGCACAGTTCACTTGGCACCACTGTGTCTTACCGGTGAC 7920
QY 2582 HisGlyPheTyrLeuLeuGlyThrProValLeuSerCysGlnIleAspGlyThrTyrPhe 2601
DB 7921 CACGGCTTCACTCTCTGAGGCAACCCAGTGTCTGAGTGCAGGAGATGAGCATGGGAC 7980
QY 2602 ArgProArgProGlnCysLeu 2608
DB 7981 CGTCCCGCCCCCAAGTGTCTC 8001

RESULT 11
ADH72215
ID ADH72215 standard; DNA; 10989 BP.
XX
AC ADH72215;
```



```
Db 1075 ATCAGTGGCAGGAGCTCCATTCATATGCTAACTGCATGACCTCCCTCCAGTT 1134
Qy 16 ILeSerAsp-----CysCysAlaSerAsnGlnArgHisSer 27
Db 1135 ATCAAGTAGAGAAAGATTGGCTAGCACTCCATTCCCTGACAGCAACCCGAGCAAA 1194
Qy 28 ValGlyValGlyProSerGluLeuValIleValysGlnIleGluLeuLysSerArgGlyVal 47
Db 1195 ---GGATTAAAGCTCAGTCCAAAGTGAAGGCGGATTGAGTTGAATCAAGAGGAGTTC 1251
Qy 48 LysLeuMetProSerLysAspAsnSerGlnLysThrSerValIleThrGlnValGlyVal 67
Db 1252 AAGATGCTGCGCCAGCAGAGGATGGAAGCATAAAACTCGTCTTGAGCCAGAGAGTCTT 1311
Qy 68 SerGlnGlyHisAsnMetCysProAspProGlyIleProGluValArgGlyLysArgLeuGly 87
Db 1312 GCATTGGTCTCACAAGTGTCTAGATCTCTGGGATTCCAGAAAATGGTAGAAGAGCAGGT 1371
Qy 88 SerAspPhe---ArgLeuGlySerSerValGlnPheThrCysAsnGlnGlyTyrAspLeu 106
Db 1372 TCCGACTTCAGTAGAGGTGGTGCAAAATGATACGTTTTCATGTGAGCAATTAAGTCTCTC 1431
Qy 107 GlnGlySerLysArgIleThrCysMetLysValSerAspMetPheAlaIleTyrSerAsp 126
Db 1432 CAGGATCTAAAGCATCACTGTCAGAGCTACAGAGAGCTCGCTGCTTGAGATGAC 1491
Qy 127 HisArgProValCysArgAlaArgMetCysAspAlaHisIleuValArgGlyProSerGlyIle 146
Db 1492 CACAGGCCCATCTGCGCAGCAGAGAACATGTGCATCTCGTGGGAGCCGAGGCGGTCC 1551
Qy 147 IleThrSerProAsnPheProIleGlnTyrAspAsnAlaHisCysValIleTyrIleIle 166
Db 1552 ATTACCTCCCTTAATTTATCCGGTTCAGATGAAATGATGACACACTGTGTGGGTGCATC 1611
Qy 167 ThrAlaLeuAsnProSerLysValIleLysLeuAlaPheGlnGluPheAspLeuGluVal 186
Db 1612 ACCACCAACGACCCGACAGAGTATCAAGCTTGCTTTGAAGAAGTTTGAGCGAGGACCA 1671
Qy 187 GlyTyrAspThrLeuThrValGlyAspGlyGlyGlnAspGlnLysThrValIleu 206
Db 1672 GGCATATACACCCCTGACGGTGGTGTGATGCTGGGAGGTGGAGACACAGATCGGTCTTG 1731
Qy 207 TyrMetSerGlnAsnAlaCysSerAspSerProHisThrProGlySerArgIleProGlu 226
Db 1732 TACGTG-----CTCAGCGGATCCAGTGTTCCTGTAC 1761
Qy 227 -----SerMetSerGlyAspIleTyrArgGlnLysThrValLeuGluIleCys 243
Db 1762 CTCATTGTGAGCATGAGCAACCAAGATGTGCTACAT----- 1797
Qy 244 ArgAspIleSerSerSerAspAlaArgSerGlySerValArgLysSerProLysThrSer 263
Db 1798 -----CTGCAGTCGAGTGTATGC----- 1815
Qy 264 AsnAlaValGluLeuValAlaProGlyIle-----ThrGluIleGlnGlnGly 278
Db 1816 -----ATTGGCTCACCCTGGGTTTAAAGCTGTTTACCAAGAAATTTAAAGGGA 1863
Qy 279 SerCysGlyAspProGlyIleProAlaTyrGlyArgArgGluGlySerArgPheHisHis 298
Db 1864 GGGTGGGGGATCTCTGAAATCCCGGCTATGGGAAGCGGAGCGGAGGAGCTTCTCAT 1923
Qy 299 GlyAspThrLeuLysPheGlnCysGlnProAlaPheGlnLeuValGlyGlnLysAlaIle 318
Db 1924 GGAAGATCACTACCTTTAAATGCCCGGCGGCTTTGAGCTGTGGGAGAGAGATTTATTC 1983
Qy 319 ThrCysGlnLysAsnAsnGlnTyrSerAlaLysLysProGlyCysValPheSerCysPhe 338
Db 1984 ACCTGTCACAGAACATCATGAGTGTCTGCAACAAGCCAGCTGTATTTTTCATGTTTC 2043
Qy 339 PheAsnPheThrSerProSerGlyValValLeuSerProAsnTyrProGluAspTyrGly 358
Db 2044 TTCACCTTACGGCATCATCTGGATTAATTTCTGTCAACCAATTAATCCAGAGAAATATGGG 2103
Qy 359 AsnHisIleuHisCysValIleTyrLeuIleLeuAlaArgProGluSerArgIleHisIleuAla 378
Db 2104 AACAACTGAACCTGTGTGTGTGATTAATCTCGAGCCAGGAAGTGAATTCACCTTAATTC 2163
Qy 379 PheAsnAspIleAspValGluProGlnPheAspPheLeuValIleLysAspGlyAlaThr 398
Db 2164 TTTAATGAATTTATGATGTGAGCTCAATTGACTTTCTGCGGCTCAAGATTAATGCAATT 2223
Qy 399 AlaGluAlaProValLeuGluLysThrPheSerGlyAsnGlnLeuProSerSerIleThrSer 418
Db 2224 TCTGACATAACTGTCTGTGGTACTTTTCTTGCAATGAAGTCCCTTCCAGCTGGCCAGC 2283
Qy 419 SerGlyHisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLysArgGlyPhe 438
Db 2284 AGTGGGCAATATAGTTGCTTGCTTGGAATTTCACTGTGACCATTTCCACTACGAGAGGTTTC 2343
Qy 439 AsnIleThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsn 458
Db 2344 AACATCACTTACACACCATTTGGTCAGATGAGTCCCATGATCTCGGCAATCTTATTAAC 2403
Qy 459 GlyLysArgPheGlyAspSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGlu 478
Db 2404 GGAACACGTTTGTGTGACAGCTTTTACTCGGAGAGCTGGTTCCTTTCACATGTATGAT 2463
Qy 479 GlyPheLeuGlyTyrGlnGlySerGluThrIleThrCysValIleuValGluGlySerVal 498
Db 2464 GGCCTTTGTCAAGACCCAGGATCCGAGTCCATTAATCTGCATATGCAAGAGCGGAACCTG 2523
Qy 499 ValTyrAsnSerAlaValLeuValCysGlnLysProCysGlyGlyHisIleuThrSerPro 518
Db 2524 GCTCGAGCTCAACCGTCCCGGCTGTGAACCTCATGTGTGACACTGTGACAGCGGCC 2583
Qy 519 SerGlyThrIleLeuSerProGlyTyrProGlyLysPheTyrLysAspAlaLeuSerCysAla 538
Db 2584 AGCGGAGTCAATTTGGCTCCTGGATGGCCAGGATTAATTAAGATTTCTTTCATTTGCA 2643
Qy 539 TyrValIleGlnValAlaGlnProGlyTyrProIleLysIleThrPheAspArgPheLysThr 558
Db 2644 TGGATTAATGAAGAAACCAAGCGCACTCATCAAAATGACTTTTGAACAGATTCACAGCA 2703
Qy 559 GluValAsnTyrAspThrLeuGluValArgAspGlyArgThrTyrSerAlaProLeuIle 578
Db 2704 GAGGTCAATTAATGACACTTGTGAGAGTCAAGATGGCCAGCCAGTTGTCCCATATC 2763
Qy 579 GlyValTyrHisGlyTyrGlnValProGlnPheLeuIleSerThrSerAsnTyrLeuTyr 598
Db 2764 GGCAGATACCAACGCGCACCCAGGACCCAGTTCCTCATGACACCGGGAATTCATGTAC 2823
Qy 599 LeuLeuPheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuAlaGlyTyrGluThr 618
Db 2824 CTGCTGTTCACACTGACCAACAGCCGCTCCAGCATGGCTTCTCATCACTATGAGAGT 2883
Qy 619 IleThrLeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlyGlnArgHis 638
Db 2884 GTGACGCTTGAATGAGATTCCTGCTGAGACCCGGGATCCCTGTGAACGCCATCCCCAC 2943
Qy 639 GlyAsnAspPheTyrValGlyAlaLeuValThrPheSerCysAspSerGlyTyrThrLeu 658
Db 2944 GGTGGAACCTTGTGCAAGTGTACAGATGACATTTTACAGTGTGACCCGGGGTACACATA 3003
Qy 659 SerAspGlyLysProLeuGlnCysGluProAsnPheGlnTyrSerArgAlaLeuProSer 678
Db 3004 AGTGACAGAGAGCCCTCGTGTGTGTGAGAGAACCACTGGAACACCCCTTGCCAC 3063
Qy 679 CysGlnAlaLeuCysGlyGlyPheIleGlnGlySerSerGlyTyrIleLeuSerProGly 698
Db 3064 TGGAGAGCTTATGATGAGAGCTAATCCAAAGGAAGAGTGAACAGTCTTCTCTCGGG 3123
Qy 699 PheProAspPheTyrProAsnAsnLeuAsnCysThrTyrIleIleGluThrSerHisGly 718
Db 3124 TTTCCAGATTTTATTCAAACTCTTAACCTGACGTTGACCATTTGAAGTGTCTCATGGG 3183
```

QY 719 LysGlyVal1PhePheThrPheHisLeuGluSerGlyHisAspTyrLeuLeu 738
DB 3184 AAGAGATTCAATATATCTTTACACCTTTGATCTTGAGAGTCCCAAGCATATTTACTG 3243
QY 739 IleThrGluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuPro 758
DB 3244 ATCAGAGAGATGAGATGTTTTTCCGAGCCCGTTGCGAGCTCACCGGGTGGTGGCTT 3303
QY 759 AlaProIleSerAlaGlyLeuThrGlyAsnPheThrAlaGlnValArgPheIleSerAsp 778
DB 3304 CATACGATCAAGCAGGCGCTGTTGAAACTTCACTGCCAGCTTGGTTATATCAAC 3363
QY 779 PheSerMetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCys 798
DB 3364 TTCTCAATTTGCTGAGAGGCGCTTCAATATCATTTTCAAGATATGACTGAGACCAT 3423
QY 799 GluGluProGluValProAlaTyrSerIleArgGlyLeuGlnPheGlyValGlyAsp 818
DB 3424 GATGATCTCGAGTCCCTGCTTCCGCGAAGAAATGGTTTTCATTGGTGGGAGAC 3483
QY 819 ThrLeuThrPheSerCysPheProGlyTyrArgLeuGluGlyThrAlaArgIleThrCys 838
DB 3484 TCTCGAGCTTTCTCTGCTTCTCGGATATCGTTTGAAGGTGCCACCAAGCTTACCTGC 3543
QY 839 LeuGlyGlyArgArgLeuThrSerSerProLeuProArgCysValAlaGluCysGly 858
DB 3544 CTGGGTGGGGCGCCCGCTGCTGAGAGTCACTCTGCCAAGGTGTGGCCGAATGTGA 3603
QY 859 AsnSerValIleThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsn 878
DB 3604 GCAAGTGTCAAGAGAAATGAAGAAATTAATGCTTCCAAATTTTCCATTCACATTAAT 3663
QY 879 AsnAsnHisGlyCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGlnLeuLys 898
DB 3664 AATACCCATGAGTATCTATTAATAAGAAAGAGCGGCAAGGCGATCACCTTGA 3723
QY 899 AlaArgAlaPheGluLeuSerGlyGlyValAspValIleuLysValTyrAspGlyAsnAsn 918
DB 3724 ACAGAGACTTCCAGCTTTTGAAGAGATCTTAAGAGTATGATGAGAAAGACAGT 3783
QY 919 SerAlaArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsnSer 938
DB 3784 TCCCTACGCTCCATGGGCGACGTTCTTAATAATGAATCTTGGGGCTGATCTTAACGC 3843
QY 939 ThrSerSerSerLeuThrLeuAspPheIleThrAspAlaGluAsnThrSerGlyPhe 958
DB 3844 ACATCCAAATCACTATGGCTAGAGTTCAACCAATGATGATCTGACACCGAACCAAGTTT 3903
QY 959 GluLeuHisPheSerSerPheGluLeuIleLysCysGluAspProGlyThrProLysPhe 978
DB 3904 CAATCACTATACCACTTTGATCTGTTAAATGTGAGATCCGGGCGATCCCTAATAC 3963
QY 979 GlyTyrIleValHisAspGlyGlyHisPheAlaGlySerSerValSerPheSerCysAsp 998
DB 3964 GGCTATAGAGTCCGATGAAGGCCACTTTCACGCACTGTATGTTCTGTAAGTTGCAAC 4023
QY 999 ProGlyTyrSerLeuArgGlySerGluGluLeuLysCysLeuSerGlyGluArgThr 1018
DB 4024 CCGGGGATACGCGATGATGAGCAACACCTGACCTGTTGAGTGGAGAGAGAGTGT 4083
QY 1019 TyrAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluVal 1038
DB 4084 TGGAGCAACCACTAATCTTCTGCTGATAGCGGAATGTGTGATCGATTCATGACCCACA 4143
QY 1039 SerGlyGlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIle 1058
DB 4144 TCAGAGCAAGATATGTCTCCCTGGCTATCCAGCTCCGATATGACAAACCTCCACTGACCC 4203
QY 1059 TyrThrIleGlyAlaGlyValGlyCysThrIleGlyLeuHisPheLeuValPheAspThr 1078
DB 4204 TGGATTATAGAGGAGACCCAGAAAGACCAATTAAGCTTCATTTGTTTTCAGACAG 4263

QY 1079 GluGluValHisAspValLeuArgIleThrAspGlyProValGluSerGlyValLeuLeu 1098
DB 4264 GAGATGGCTACGACATCTCTCAAGCTCGGAGCGGGCGGTGAGACAGTGAATCTGCTG 4323
QY 1099 LysGluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerVal 1118
DB 4324 AAGAGTGAATGAGCTCCGCTTCCGGAGAGATCCACAGACCTTCACTCACTCAC 4383
QY 1119 LeuGlnPheSerThrAspPheThrPheThrSerIleGlnGlyPheAlaIleGlnPheSerVal 1138
DB 4384 CTGAGATTGACACAGACTTTTTCATCAAGAACTGTGCTTCTCATTCACATCTTCCACC 4443
QY 1139 SerThrAlaThrSerCysAsnAspProGlyIleProGlnAsnGlySerArgSerGlyAsp 1158
DB 4444 TCAATTGAGGACCACTGTGAAGATCCAGATGCCCCAATAATGCAACCCGTATGAGAC 4503
QY 1159 SerTyrGluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGly 1178
DB 4504 AGCAGAGAGCTGAGACACCGTCAATTCAGTGTGACCTGGCTATCACTCCAAAGA 4563
QY 1179 SerAlaGluIleSerCysValIleGlyIleGluAsnArgPheThrGlnProSerProPro 1198
DB 4564 CAAGCCAAATCACTGTGTGACAGTGAATTAACCGGTTCTTTGGCAACCAAGACCTTCT 4623
QY 1199 ThrCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerPro 1218
DB 4624 ACATGCAATAGCTGTTGTGAGGAAATCTGACGGGCCCAAGAGTATTTTGTACCC 4683
QY 1219 AsnTyrProGluProTyrProProGlyValGlyGlyCysAspTyrIleValSerPro 1238
DB 4684 AACTACCCACAGCGGTATCTCTCGGAGAGATGTGATGAGAGATTAAGTAAGTAACCCG 4743
QY 1239 AspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeu 1258
DB 4744 GACTTTGTCACTCGCTTGATTAATTAACAAATGTTCAATGAGGCCACATGATGATCTTA 4803
QY 1259 HisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGln 1278
DB 4804 CACATCTATGAGGGAGAGATTTCCACAGCCCTTCTTGGAGATTACAGAGGCTTCCAG 4863
QY 1279 LeuProGlyValIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAla 1298
DB 4864 GCCCCAGAAAGATAGAGAGTGAAGGAGAAACGCTGTTTCTGGCATTTCCGAGTATGCC 4923
QY 1299 SerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPhe 1318
DB 4924 TCCGTGGGCTTTTCAGGGTTCCGCAATGAATTTAAAGAAACCAAGGAAAGCTTGT 4983
QY 1319 AspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySerSer 1338
DB 4984 GACCCAGAAATATATGAATGGAGCAAGAGTTGGAACAGACTTCAAGCTTGGCTCCACC 5043
QY 1339 ValThrTyrTyrCysHisGlyGlyIleGluValGluGlyThrSerThrLeuSerCysIle 1358
DB 5044 ATCACTACCAAGTGTGACTCTGGCTATGAATTTCTGACCCCTCATCATCAGCTGTGTG 5103
QY 1359 LeuGlyProAspGlyLysProValTyrAsnAsnProArgProValCysThrAlaProCys 1378
DB 5104 ATTTGGGCTATAGGGAACCTCTCTGGGACCAAGTGTGCTTCTCTCAATGCTCCCTGT 5163
QY 1379 GlyGlyGlnTyrValGlySerSerAspGlyValValLeuSerProAsnTyrProGlnAsnTyr 1398
DB 5164 GAGGCGCAAGTACAGGAGTACGAAGGGGTAGTTTATCAACCAATCAACCCCATATTAAC 5223
QY 1399 ThrSerGlyGlnIleCysLeuTyrPheValThrValProLysAspTyrValAlaPheGly 1418
DB 5224 ACAGCTGTCAATATGCTCTATTCATCAACCGTACCAAGAAATTCGTGTGCTTTGGA 5283
QY 1419 GluPheAlaPhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSer 1438
DB 5284 CAGTTTGCCTATTTTCCAGACAGCCCTGAAGATTTGCAAGAAATTAATTTGATGGAACCAT 5343
QY 1439 GlnHisSerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeu 1458

```
Db 5344 GCACAGCCAGACTTCTACGCTCAGCTCGGGGTCTACTCGAGGAAACATGCTCTTG 5403
Qy 1459 AATATSerAsnGlnValLeuIleLysPheSerAlaLysGlyLeuAlaProAlaArgGly 1478
Db 5404 GGTAGCTCAAAATCAATTTCTGCTCGATTCAGTGCAGAAAGCGGTGCTTGCCGGCGGC 5463
Qy 1479 PheHisPheValTyrGlnAlaValProArgThrSerAlaThrGlnCysSerSerValPro 1498
Db 5464 TTCACCTTGCTATACAGCTTCTCTGACAGTACACCCCAATGACGCTGTCCTCC 5523
Qy 1499 GlnProArgTyrGlyLysArgLeuGlySerAspPheSerValGlyAlaIleValArgPhe 1518
Db 5524 GAGCCAGATACGAGAAAGAAATGCTTCTGATTTCTGCGCGGTCAATGCTCCGATTC 5583
Qy 1519 GlnCysAsnSerGlyTyrAlaLeuGlnGlySerProGluIleGluCysLeuProValPro 1538
Db 5584 GAGTCAACCGGGGATACCTGCTTCAGGGTTCACGGCGCTACCTGACCATTCGCGTCCC 5643
Qy 1539 GlnAlaLeuAlaGlnTyrAsnValSerAlaProThrCysValValProCysGlyGlyAsn 1558
Db 5644 AACGCTTGCGACAGTGCAGACAGACAGATCCCGAGCTGTGTGTACCTGACATGGCAAT 5703
Qy 1559 LeuThrGlnArgArgGlyThrIleLeuSerProGlyPheProGluProTyrLeuAsnSer 1578
Db 5704 TTCACTCACAGAGAGGTACAAATCCTGCCCCGCTACCTGAGCCATACGAGAAACAC 5763
Qy 1579 LeuAsnCysValTyrPheIleValValProGlnGlyAlaGlyIleGlnIleGlnValIle 1598
Db 5764 TTGAAGCTATATAGAGATCATAGTACGAGAGGCTCGGGAATTCAGAGATTCAGAGAT 5823
Qy 1598 IserPheValThrGlnGlnAsnTyrAsp-SerLeu-GluValPheAspGlyAlaAspAsn 1617
Db 5824 CAGTTTGCCACGAGACAGAACTGGAATCCCTTACAGATCCAGATGGTGGGATGTG 5883
Qy 1618 ThrAlaThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThr 1637
Db 5884 ACCGACCCGACAGCTGGAGAGCTTCTCAGGACACCACTACCGGACACTGGAACAGATAC 5943
Qy 1638 SerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSerValSerAlaIleGlyPheHis 1657
Db 5944 TCCAAACCACTACCTGACATTTCCAGTCTGACATTTAGTGGAGCTGTGGTTCCTCC 6003
Qy 1658 LeuGluTyrIleThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGly 1677
Db 6004 CTGGAATAACAAACTGTAGGCTTGTCTGATGCCAAGAACCAAGCCCTCCCAACACAC 6063
Qy 1678 ValIleThrGlnGlyLysArgTyrLeuValAsnAspValValSerPheGlnCysGluProGly 1697
Db 6064 ATCAAAATCGAAGTCCGATCATGTGACGACGAGCTGCTCTCCACAGTCCGAGCCGGG 6123
Qy 1698 TyrAlaLeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTyrAsn 1717
Db 6124 TACACCTGCAGGCGCGTCCCACTTCTCTGATGCGAGGACCTTGCCGCTTGAGAC 6183
Qy 1718 TyrProProProLeuCysIleAlaGlnCysGlyGlyThrValGlnGluMetGlnGlyVal 1737
Db 6184 TATCCGCTTCCCTGTGCATTCGAACCTGTGAGGAGACCTGTGAGCACTTGAGGTGTG 6243
Qy 1738 IleLeuSerProGlyPheProGlyAsnTyrProSerAsnMetAspCysSerTrpIle 1757
Db 6244 ATCTAGAGCCCGGCTTCCAGGTTCTTACCCCAACACTGACCTGAGAGATC 6303
Qy 1758 AlaLeuProValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsn 1777
Db 6304 TCATTTCCCATCGGCTATGTGACATATTCAGTTCTGATTTTCTACCGAAGCTAAT 6363
Qy 1778 HisAspTyrIleGlnIleArgAsnGlyProTyrGluThrSerArgMetCysArgPhe 1797
Db 6364 CATGACTTCTTGAATTCAGAAATGAGCTTACCAACACCGCCCATGATGAGCAATTT 6423
Qy 1798 SerGlySerGluLeuProSerSerLeuSerThrSerHisGluThrValIleYrPhe 1817
Db 6424 AGCGGACAGGATCTCCCGGCGCTGTGAGACACAACCATGAAACCTCATCACTTT 6483
Qy 1818 HisSerAspHisSerGlnAsnArgProGlyPheLysLeuGluTyrGlnAlaTyrGluLeu 1837
Db 6484 TATAGTACCATTTCCGAAACCGGACAGAAATTTAACTTTCACCAACCTATGAAATTA 6543
Qy 1838 GlnGluCysProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrAsn 1857
Db 6544 CAGAACTGTCCAGATCACCCCATTTGAGATGGGTATCATATGCACTCGGATTACAC 6603
Qy 1858 ValGlyGlnSerValThrPheGluCysLeuProGlyTyrGlnLeuThrGlyHisProVal 1877
Db 6604 GTGGGCAATCATATCTTTCAGATGTATCTGGGTATCATTTCAATGAGCATCTGTTC 6663
Qy 1878 LeuThrCysGlnHisGlyThrAsnArgAsnTyrAspHisProLeuProLysCysGluVal 1897
Db 6664 CTCACCTTGACAGATGGGATTCACAGAAATGMACTGACCTTTCCAAAGATGTAGTGC 6723
Qy 1898 ProCysGlyGlyAsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSer 1917
Db 6724 CCTTGGGTACAGATTAATCTTCAGAAACGACACATCTACTCCCTGGCTTCTGAT 6783
Qy 1918 ProTyrSerSerSerGlnAspCysValTrpLeuIleThrValProIleGlyHisGlyVal 1937
Db 6784 GAGTATCCGATCTGAAGAGCTGCAATTTGGCTCATCAGGTGCTCCAGGCGCAGAGTT 6843
Qy 1938 ArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAsp 1957
Db 6844 TACATCAACTTCAACCTGTTACAGACGAGGTGCACAGATTCATGCTGTTGGAC 6903
Qy 1958 GlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysLeuThr 1977
Db 6904 GGTCCGATTCAGAACTCAACCCAGCTGGAGTTTTCAGTGCACACAGCCCTCCAAAGC 6963
Qy 1978 ValGlnSerSerSerAsnGlnValLeuLeuLysPheHisArgAspAlaIleThrGlyGly 1997
Db 6964 GCGTATACCTCACCACCAAGTCTCTGTCAGTTCCACAGGACTTTTCAAAATGAGAGC 7023
Qy 1998 IlePheAlaIleAlaPheSerAlaTyrProLeuThrIleCysProProProThrIleLeu 2017
Db 7024 TTCCTTGTCCCAATTTCCACGCAATTCAGCTCAAGAAATGTCAACCTCCCGAGCGGTT 7083
Qy 2018 ProAsnAlaGluValAlaThrGlnAsnGlnGluPheAsnIleGlyAspIleValArgTyr 2037
Db 7084 CCAAGGCAAAATGCTTACTAGAGATGATATTTCAAAATGAGAGATTTTGTGAAGTAC 7143
Qy 2038 ArgCysLeuProGlyPheThrLeuValGlyAsnGlnIleLeuThrCysLysLeuGlyThr 2057
Db 7144 CAGTGCACCCCGGTTACACTTGTGGGACCGACATTTGACTTGCAGCTTCAGTTC 7203
Qy 2058 TyrLeuGlnPheGlnGlyProProProIleCysGluValHisCysProThrAsnGluLeu 2077
Db 7204 CAGTTGCAATTTGAGGTTCTCTCCCAATGTGMAACACAATGCCAGCAAAATGAGATC 7263
Qy 2078 LeuThrAspSerThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPhe 2097
Db 7264 CGAGCTGATATCGGAGATCATTTCTCAGTCCAGGAGTATCCGGTATTTTAACTCC 7233
Qy 2098 GlnThrCysSerTyrPheValArgValGluProAspTyrAsnIleSerLeuThrValGlu 2117
Db 7234 CAGACTTGCTTGGAGTATTTAAAGTGAACCAAACTACAAACATTCACATCTTGTGGAC 7283
Qy 2118 TyrPheLeuSerGlnLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGln 2137
Db 7284 ACATTTCAAGTGAAGAGAGTTTGAAGCACTGAAAGTGTATTAAGTCTTCTGGGCA 7443
Qy 2138 SerProLeuLeuLysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSer 2157
Db 7444 AGTCCCTGTAGAGTCTTAAGTGGAAATCATACGAAACATCAATTTTACAGAGAG 7503
Qy 2158 SerAsnSerValTyrLeuArgTyrPserSerAspHisAlaTyrAsnArgIleGlyPheLys 2177
Db 7504 AGTAAATCAATTAATCTCCGCTGTGCTCACTGACCATGACCAAGTGAAGAGATTCAG 7563
```

OY	2538	SeArpMeLeuProThrCysArgIleLeuLenCysThrAspProGlyIleGlnGluAsn	2557
Dp	8644	AGTACCCCTCTGCCACGCTGTCAGTGTGAAGCTGCTATCCAGGCTTTGTGGAAAT	8703
OY	2558	SeValAlaArgIleValIleAlaSerGlyProHisArgPheSerPheGlyThrThrValSer	2577
Dp	8704	GCACATTTGTCACGGGCAACAGAACTCCCTGAGAGTTTGTGATGGAAATGATATCCTG	8766
OY	2578	TyrArgCysAsnHisGlyPheThrGlyLeuGluGlyThrProValLeuSerCysGlnGlyAsp	2597
Dp	8764	TACCATTTGCAGAAAGGAGATTTTACTGTGTGGATCTTCAGCCTTCAGCCTGATATGCAAAAT	8823
OY	2598	GlyThrThrPheAspArgProArgProGlnCysLeuLeuValSerCysGlyIleAspProGlySer	2617
Dp	8824	GGCTTATATGGACCCATCTCCCTGCCAAGTGTGTGCTATATGCTGTGGACACCCAGGGCT	8883
OY	2618	ProPheHisSerGlnMetSerGlyAspSerTyrThrValGlyAlaValAlaArgTyrSer	2637
Dp	8884	CTGCGCAACGGCGGCTCCTACGTCGAGAGCTGTTTACATATGGCGCGCTGTGCATATCCTCC	8943
OY	2638	CysIleGlyIleArgThrIleValGlyAsnSerThrArgMetCysGlyLeuAspGlyHis	2657
Dp	8944	TGCAGAGGGAGCGAGAGCTCATAGGCAACGACACGAGAGTGTCCAGGAAGACATCAC	9003
OY	2658	ThrPheArgIleSerLeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGly	2677
Dp	9004	TGGAGCGGGGACATCGCCCATGTGCAGAAATATCTGATTTCTGTGGTATCCGGGG	9063
OY	2678	IleProAlaHisGlyIleArgLeuGlnYAspSerPheAspProGlyThrValMetArgPhe	2697
Dp	9064	ACCCACAGCATGGGCTCTGGCTGTGTGATGACTTTAAGACAAAGAGCTTCTCCGCTTC	9123
OY	2698	SerCysGlnAlaGlyHisValLeuArgGlySerSerGlnArgThrCysGlnAlaAsnGly	2717
Dp	9124	TCTGTGATAAATGGGGACACAGCTAGAGGGCTCCCTGAAACGACGATGTTGCTCAATGGG	9183
OY	2718	SerThrSerArgIleSerGlnProGluCysGlyValIleSerCysGlyAspProGlyThrPro	2737
Dp	9184	TCAATGTGCAGACATGCAGCGCGGTGTGAGGGCGCTGTCTGTGGCAACCTCGGACACCC	9243
OY	2738	SerAsnAlaArgValAlaPheSerAspGlyLeuValPheSerSerIleValTyrGlu	2757
Dp	9244	ACCAACGGAAATGATTTGTCAGTACAGTATGATGGCAATTTCTTCTCCAGCTGGGTATCTATGCC	9303
OY	2758	CysArgGlnGlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyThr	2777
Dp	9304	TGCTGGGAGGCTTACAAAGACTCAGGGGTATGACACGGCAATTCACAGCAAGGGAGCC	9363
OY	2778	ThrThrArgIleSerAspProGluCysLeuValIleAsnGlyAspProGlyThrProAla	2797
Dp	9364	TGGACAGGACCTGCTCCCGGACTGCAGCACTTAATAGTTGTGGGGATTCACAGGCACATGACA	9423
OY	2798	AsnGlyLeuArgLeuGlyAsnAspPheArgTyrThrAsnIleThrValIleTyrGlnCysVal	2817
Dp	9424	AATGGCAATCCAGTTTGGACCGGACTTCACTTCAACAAGAGCTGAGCTATCAAGTATAC	9483
OY	2818	ProGlyTyrMetMetGlnSerHisArgValSerValLeuLeuSerCysThrIleAspArgThr	2837
Dp	9484	CCAAGCTATGTCATGAGAAAGCATACATCCGCGCATATGCGCTGATCCAAAGAAGCGGACG	9543
OY	2838	ThrPheArgIleThrIleAspProValCysValAlaLeuMetCysIleAspProProLeuIlePro	2857
Dp	9544	TGGAAATCCGAGCAAACTGCTCGCAAAAGCCGCTGTGTCTCTCAAGCCGCGCCGCTGACG	9603
OY	2858	AsnGlyIleValValGlySerAspPheMetThrGlySerSerValIleTyrAlaCysLeu	2877
Dp	9604	AATGGAACAGTGGAGGAAAGTATTCGCTGGGGCTCCACGATTAAGCTACAGCTGATG	9663
OY	2878	GlyGlyIleTyrGlnLeuSerLeuProAlaValPheThrCysGlnGlyIleAsnGlySerThrThr	2897
Dp	9664	GACGGTTATACAGAGCTCTCTCACTCCGCGCATATCTCTCCGTAAAGTCCGGGGGTGGAAA	9723
OY	2898	GlyIleLeuLeuProGlnCysPheProValPheCysGlyAspProGlyValProSerArgGly	2917

```

Db      9724 GGAGAGATCCCCAGTGTCTCGCTGTCTTCTGCGAGACCTTGCATCTCCGCGAAGGG 9783
Qy      2918 ArgArgGluAerArgGlyPheSerTyArgSerSerValSerPheSerCysHisProPro 2937
Db      9784 GCACCTTAGTGGGAAAAGTTTCCACTATTAAGTCCGAAGTCTTCTTCCATGCAAAATCTCCA 9843
Qy      2938 LeuValLeuValGlySerProArgArgPheCysGlnSerArgProGlyTyrTrpSerGlyThr 2957
Db      9844 TTTATCTCGTGGATCTCTCCAGAAAGTCTGCGAAGCTGACGACGACGTGAGCGGACATA 9903
Qy      2958 GlnProSerCysIle---AspProThrLeuThrThrCysAlaAspProGlyValProGln 2976
Db      9904 CAACCCACCTGATGTGATCTGATCTGCTCAACACCTGCCAGACCTCTGAGCCACAC 9963
Qy      2977 PheGlyIleGlnAerAerSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCys 2996
Db      9964 TTGGAAATACGAATAGCTCCAGAGGCTATGAGGTGGAGACAGGTTTTCAGGTGC 10023
Qy      2997 GlnArgGlyTyrLeuLeuGlnGlySerThrThrArgThrCysLeuProAerLeuThrTrp 3016
Db      10024 AGAAAGGCTACCATTTCAAGGTTTCCAGACCTGCGACCTGCTTGCCTCAATTAACTGG 10083
Qy      3017 SerGlyThrProProAspCysValProHisHisCysArgGlnProGlyThrProThrHis 3036
Db      10084 AGTGGATACAGACCGAATGATATGATCATGCTGCGACAGACGCAAAACCCGGGCACAC 10143
Qy      3037 AlaAerValGlyAlaLeuAerLeuProSerMetClyTyrThrLeuIle-ThrProAlaAr 3056
Db      10144 GCGAGTGTGAGGACCTGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 10203
Qy      3056 GArgAlaSerProSerArgValAlaProSerThrAlaProAlaArgArgMetAlaIle 3076
Db      10204 CCAGGCTTTTCTCTGCGAGGGGATCTGAGACAGAACATGTAAGCAGACATGAATGG 10263
Qy      3076 YGlnAlaSerArg 3080
Db      10264 ACAGGAAAGTGC 10276

RESULT 12
AAD33320
ID AAD33320 standard; cDNA; 10433 BP.
XX
AC AAD33320;
XX
01-JUL-2002 (first entry)
XX
Human C3b/C4b complement receptor like cDNA #2.
DE
XX
Human C3b/C4b complement receptor like cDNA #2.
XX
Human; C3b/C4b complement receptor-like molecule; immune system disorder;
KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;
KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
KW transplant rejection; autoimmune disease; ischaemic condition; noctropic;
KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
KW infertility; vasodilator; obesity; cardiac; gene; ss.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 1. 9903
FT /tag= a
FT /product= "Human C3b/C4b complement receptor like protein
FT #2"
FT /note= "CDS does not include start codon"
FT /transl_except= (pos:2176..2178, aa:Xaa)
FT /note= "This translational exception occurs while
FT decoding the alternative version of human C3b/C4b
FT complement receptor like protein #2 (AAE20901)"
FT /partial
XX
XX WO200210199-A2.
XX

```

```

PD      07-FEB-2002.
XX
XX      24-JUL-2001; 2001WO-US023232.
PF
XX      02-AUG-2000; 2000US-0222504P.
PR
XX      28-NOV-2000; 2000US-00728787.
XX
XX      (AMGE-) AMGEN INC.
XX
XX      Welcher AA, Elliott GS;
XX
XX      WPI; 2002-303934/34.
DR
XX      P-PSDB; AAE20789, AAE20901.
XX
XX      Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
PT sclerosis.
XX
XX      Claim 1, Fig 2; 251p; English.
XX
XX      The invention relates to a nucleic acid encoding a novel C3b/C4b
CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
CC polypeptide and nucleic acid molecules may be used to treat, prevent,
CC ameliorate, diagnose and/or detect diseases such as immune system
CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
CC transplant rejection, nervous system disorders (e.g. Alzheimer's
CC disease), ischaemic conditions, metabolic disorders (e.g. obesity and
CC diabetes) and infertility. The invention is useful in gene therapy. The
XX present sequence is human C3b/C4b complement receptor like cDNA
XX
XX      Sequence 10433 BP; 2618 A; 2671 C; 2527 G; 2610 T; 0 U; 7 Other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 10433
Score: 10760.00 Matches: 1865
Percent Similarity: 79.24% Conservative: 433
Best Local Similarity: 64.31% Mismatches: 557
Query Match: 63.47% Indels: 45
DB: 6 Gaps: 5
US-10-016-248-2 (1-3104) x AAD33320 (1-10433)
Qy      190 ThrLeuThrValGlyAerGlyGlyGlnAerGlyAerGlnThrValLeuThrMetSer 209
Db      1 ACCCTGACGGTGTGATGCTGCGAAGTGGAGACACAGATCGCTTGTACGTG--- 57
Qy      210 GlnAerAlaCysSerAerSerProHisThrProGlySerArgIleProGlu----- 226
Db      58 -----CTACGGGACCTCAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 226
Qy      227 SerMetSerGlyAerIleTyrArgGlnIleTyrThrValLeuGlnIleCysArgAspIle 246
Db      91 AGCATGAGCAACACGATGTGGCTAAT-----CTG 120
Qy      247 SerSerSerAerAlaAerSerGlySerValArgLysSerProLysThrSerAerAlaVal 266
Db      121 CAGTGGATGATAC----- 135
Qy      267 GluLeuValAlaProGly-----ThrGlnIleGlnGlnIleGlnIleGlnIleGlnIle 281
Db      136 ---ATTGGCTCACCTGGGTTTAAAGCTGTTTACCAAGAAATTAAGAGGAGGTGG 192
Qy      282 AspProGlyIleProAlaTyrGlyArgArgGlnGlySerArgPheHisGlyAerThr 301
Db      193 GATCTTGAAATCCCGGCTATGGGAAGCGGACGAGAGTTCCTCATGAGATACA 252
Qy      302 LeuLysPheGlnGlyGlnProAlaPheGlnLeuValGlyGlnAlaIleThrCysGln 321
Db      253 CTCACCTTGAATAGCCGCGGCGCTTGTGAGCTGTGGGAGAGAGTATACCTGTCA 312

```

```
QY 322 LysAnaEngInTPSeAlaLeYsProGlyCyValPheSerCyPhePheAspPhe 341
Db 313 CAGAAACATCAGTGGTCTGGCAACAGCCAGCTGTGATTTTCACTTTCTTCACTTT 372
QY 342 ThiserProSerGlyValValLeuSerProAnTYrProGluAspTYrGlyAsnHisLeu 361
Db 373 ACGGCATCATCTGGGATTAATCTGTGACCAATATACAGAGAAATATGGGAACAACATG 432
QY 362 HisCyValATTPLeuLleLeuAlaArgProGluSerArgLleHisLeuAlaPheAsnAsp 381
Db 433 AACGTGTCTGTGATTAATCTCGAGAGCCAGAAATCGAATTCACCTTAATCTTAATCAT 492
QY 382 IleAspValGluProGlnPheAspPheLeuValIleLysAspGlyAlaThrAlaGluAla 401
Db 493 TTTGATGTGAGCTCTCAATTTGACTTTCTCCGGGTCAAGATGATGGCATTTCTGACATA 552
QY 402 ProValLeuGlyYThrPheSerGlyAsnGlnLeuProSerSerLleThrSerSerGlyHis 421
Db 553 ACTGTCTGGGTACTTTCTTGTGGCAATGAGTGCCTTCCAGCTGGCCAGAGTGGGAT 612
QY 422 ValAlaArgLeuGluPheGlnThrAspHisSerThrGlyValArgGlyPheAsnLleThr 441
Db 613 ATAATGTGCTTGAAATTCAGTGCACCAATTCACACTGCGACAGAGGTTCAACATCACT 672
QY 442 PheThrThrPheArgHisAsnGluCyAspProAspProGlyValProValAsnGlyLysArg 461
Db 673 TACACCACTTTGTCAGAAATGAGTGCATATCTGTGACATTTCTATTAACAGGACGAGCT 732
QY 462 PheGlyAspSerLeuGlnLeuGlySerSerLleSerPheLeuCyAspGluGlyPheLeu 481
Db 733 TTTGTGTGACAGGTTTCTACTCGGGAGCTCGGTTTCTTTCCACTGTGATGATGGCTTGTCTC 792
QY 482 GlyYThrGlnGlySerGluThrLleThrCyValLeuLysGluGlySerValAlaTrpAsn 501
Db 793 AAGACCAAGGATCCGAGTCCATTAACCTGACTGCAAGACGGGAACGTGTGTGAGC 852
QY 502 SerAlaValLeuArgCyGluAlaProCyGlyGlyHisLeuThrSerProSerGlyYThr 521
Db 853 TTCACCGTGGCCCCGTGTGAAAGCTCCATGTGTGACATCTGACAGCTCCAGCGGAGCTC 912
QY 522 IleLeuSerProGlyYThrProGlyPheYrLysAspAlaLeuSerCyAspAlaTrpValIle 541
Db 913 ATTTGTGCTCTGTGATGGCCAGGATATATTAAGATTTCTTAACATTTGAAATGATTAAT 972
QY 542 GluAlaGlnProGlyYrProIleLysLleThrPheAspArgPheLysThrGluValAsn 561
Db 973 GAACCAAAACCAAGGACCTCTATCAAAATACTTTGACGATTTGACAGAGTCAAT 1032
QY 562 TyrAspThrLeuGluValArgAspGlyArgThrYrSerAlaProLeuLleGlyValYr 581
Db 1033 TATACACCTTTGAGGTGAGATGAGGAGCCAGCCAGTGTGTCCCACTGATCGGAGATAC 1092
QY 582 HisGlyYThrGlnValProGlnPheLeuLleSerThrSerAsnTYrLeuYrLeuLeuPhe 601
Db 1093 CACGGACACCCAGGACCCCAAGTCTCTATCAGACCCGGGAATTCAGTACTCTCTATTC 1152
QY 602 SerThrAspLysSerHisSerAspLleGlyPheGlnLeuArgYrGluThrLleThrLeu 621
Db 1153 ACCACTAGCAACACCGCTCCAGATCGGCTTCTCTCACTTCACATTTGAGAGTACACCTT 1212
QY 622 GlnSerAspHisCyLeuAspProGlyYlLeProValAsnGlyLysArgHisGlyAsnAsp 641
Db 1213 GAGTCTGATTTCTCTGTGAGCCCGGAGATCCCTGTGAAACGRCATTCGCGACAGGAGAC 1272
QY 642 PheYrValGlyAlaLeuValThrPheSerCyAspSerGlyYrYrThrLeuSerAspGly 661
Db 1273 TTTGGCATCAGGTCCACAGTACTTTCACTGATGACCCGGGTACACACTAAGAGAGAC 1332
QY 662 GluProLeuGluGlyGluProAsnPheGlnTrpSerArgAlaLeuProSerCyGluAla 681
Db 1333 GAGCCCTCTGTGTGAGAGAACACACAGTGAACACACGCTTGCCAGCTGCGAGAGCT 1392
QY 682 LeuCyGlyGlyPheLleGlnGlySerSerGlyYrThrLeuSerProGlyPheProAsp 701
Db 1393 CTATGTGAGGCTACATCCAGGAAAGATGGAACATCTTCTCTCGGTTCCAGAT 1452
QY 702 PheYrProAsnLeuAsnCyThrTrpLleGlnThrSerHisGlyLysGlyVal 721
Db 1453 TTTTATCAAACTCTTAACACGCACTGAGCCATTAAGATGTCTCATGGGAAAGAGT 1512
QY 722 PhePheThrPheHisThrPheHisLeuGlnSerGlyHisAspTYrLeuLeuLleThrGlu 741
Db 1513 CAATGATCTTTCACACCTTTCATCTTGAAAGTTCACAGACTTAATTAATGATCAGAG 1572
QY 742 AsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaProIle 761
Db 1573 GATGAGATTTTCCAGACCCCGTGGCCAGGCTCACCGGGTCCGTTGCTCATACATC 1632
QY 762 SerAlaGlyLeuYrGlyAsnPheThrAlaGlnValArgPheLleSerAspPheSerMet 781
Db 1633 AAGCAGAGCTCTTGGAACCTTCACTGCCAGGTTGCGTTTATACAGACTTCTCAAT 1692
QY 782 SerYrGluGlyPheAsnLleThrPheSerGluYrYrAspLeuGluProCyGluGluPro 801
Db 1693 TCGTACAGAGGCTTCAATATCACATTTTCAGAAATATGACCTGAGCCATGTATGATCTT 1752
QY 802 GluValProAlaYrSerLleArgLysGlyLeuGlnPheGlyValGlyAspThrLeuThr 821
Db 1753 GAGTCCCTGCTTACCGCCAGACCAAGATGTTTTCATCTTGTGTGGAGACTCTCTAGC 1812
QY 822 PheSerCyAspPheProGlyYrArgLeuGluGlyYThrAlaArgLleThrCyLeuGlyGly 841
Db 1813 TTTTCTGCTTCTCGGATATCGTTTGAAGGTGCCCAAGCTTACTCTGCTCGGTGGG 1872
QY 842 ArgArgArgLeuTrpSerSerProLeuProArgCyValAlaGluCyGlyAsnSerVal 861
Db 1873 GCGCGCGGTGTGAGATGACCTGTGCAAGGTGTGTGCCAATGTGAGCAAGTCTC 1932
QY 862 ThrYrYThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTYrAsnAsnHis 881
Db 1933 AAGGAATGAGGAACATTAATCTGTCTCCAAATTTTCCATCCAAATTAATTAATCAAT 1992
QY 882 GluCyLleYrSerLleGlnThrGlnProGlyLysGlyLleGlnLeuLysAlaArgAla 901
Db 1993 GAGGTATCTATTAATAATAGAAACAGAACCGGACAGGACTCCATTAAGACAGAAC 2052
QY 902 PheGluLeuSerGluGlyAspValLeuLysValYrAspGlyAsnAsnSerAlaArg 921
Db 2053 TTCACGTGTGTTGAGAGATCTCTAAGGTATATATGAAAGACACTTCTCTACGT 2112
QY 922 LeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerTrpSerSer 941
Db 2113 CCACGTGGACGTTCTACTAAATAATGAATCTTGGGCTGTATCTTAACACACATCCAAAT 2172
QY 942 SerLeuTrpLeuAspPheLleThrAspAlaGluAsnThrSerYsGlyPheGluLeuHis 961
Db 2173 CACCTRTGGCTAGGTTCAACACCAATGATCGACACGACAGGATTTTCAACTCAC 2232
QY 962 PheSerSerPheGluLeuLleLysCyGluAspProGlyYrProLysPheGlyYrYrLys 981
Db 2233 TATAACGATTTTATCTGTAATGAGAAATGAGAAATCCCGGCAATCCCTAATCGGCTAAGG 2292
QY 982 ValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCyAspProGlyYr 1001
Db 2293 ATCGTATATAGAGCCACTTACCGACACTGTGATTTGTACAGTTGCAACCCGGGATAC 2352
QY 1002 SerLeuArgGlySerGluGluLeuLeuCyLeuSerGlyGluAlaArgArgTrpAspArg 1021
Db 2353 GCCATGATGAGCAACAACCTGACCTGTTTGAAGTGAACAGAGAGTGTGGACAAA 2412
QY 1022 ProLeuProThrCyValAlaGluCyGlyGlyYrThrValArgGlyGluValSerGlyGln 1041
Db 2413 CCACTACTTTCGTCATGAGGATGTGGTGTCAATTCATCAGCCACATCAGAGCA 2472
QY 1042 ValLeuSerProGlyYrProAlaProYrGluHisAsnLeuAsnCyLleTrpThrLle 1061
Db 1061
```



```

Db      2473 ATATTGTCCTGGCTATCCAGCTCCGATGACAAACCTCCTGCTGATGATATA 2532
Qy      1062 GIUALAGLUALAGLYCether11leglyleuHisPheLeuValPheAspThrGluGluVal 1081
      2533 GAGGAGAGACCCGAGAAAGACATTAGCTCTTCCATTGATTTTGACAGAGATGGCT 2592
Qy      1082 HisAspValLeuArg11eTyrAspGlyProValGluSerGlyValLeuLeuGlyGlu 1101
      2593 CACGACATCTCAAGGTCTGGGACGGGCGGAGCACTGACATCTCTCTCAAGAGATGG 2652
Db      1102 SerGlyProAlaLeuProValAspLeuHisSerThrPheAsnSerValValLeuGlnPhe 1121
      2653 AGTGCTCCGCGCTTCGAGAGACATCCAGACACTTCACTCACTCACTCCCTCACTTC 2712
Qy      1122 SerThrAspPhePheThrSer11eGln11eGlnPheAla11eGlnPheSerValSerThrAla 1141
      2713 GACGCGACCTTCTTCATGACAGAGCTGGCTTCTTCCATCTCCAGTTCTCACTCAATTGCA 2772
Qy      1142 ThrSerCyAsnAspProGly11eProGlnAsnGlySerArgSerGlyAspSerTrpGlu 1161
      2773 GCCACCTGTAAAGATCCAGGTATGCCCCAAATGAGCCGCTATGAGAGAGAGAG 2832
Db      1162 AlaGlyAspSerThrValPheGlnCyAspProGly11eValAlaLeuGlnGlySerAlaGlu 1181
      2833 GCTGAGACACCGCTCACTTCAGGTGACCTCTGCTATCAGCTTCAAGAGCAAGCCAAA 2892
Qy      1182 11eSerCyVal11eGlnAsnArgPhePheTrpGlnProSerProProThrCys11e 1201
      2893 ATCACTGTGTGACAGATTAACCGGTCTTTTGGACACCAAGCCCTCTCACTGATATA 2952
Qy      1202 AlaProGlyGlyGlyAspLeuThrGlyProSerGlyVal11eLeuSerProAsnTrpPro 1221
      2953 GCTCTGTGTGAGGAAATCTGACGGGCCAGAGGTATTTTGTGACCACTCACTACCA 3012
Db      1222 GluProTrpProProGly11eGlnCyAspTrpValThrValSerProAspTrpVal 1241
      3013 CAGCGATATCTCTCTGGAGAGATGTGACTGAGAGATTAAGTAAACCCGAGCTTGTTC 3072
Qy      1242 11eAlaLeuValPheAsn11ePheAsnLeuGlnProGly11eTrpAspPheLeuHis11eTrp 1261
      3073 ATCCCTTGATATTAACAAGTTTCAACATGAGAGCCGATATGACTTCCATACACATCTAT 3132
Db      1262 AspGlyArgAspSerLeuSerProLeu11eGlySerPheTrpGlySerGlnLeuProGly 1281
      3133 GAGGAGGAGATTCACACAGCCCTCTCATTTGGAGATTAACAGAGGCTCTCAGGCCCAAGAA 3192
Qy      1282 Arg11eGlnSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerValSer 1301
      3193 AGATATAGAGTACGCGAAGACAGCTCTTCTGCACTTTCGAGTATGCTGCTCGTGGCC 3252
Db      1302 AsnAlaGlyPheVal11eAspTrpThrGluAsnProArgGluSerCyAspPheAspProGly 1321
      3253 CTTCAGGCTTCCCATTTGAATTTAAAGAGAAACCAAGGAGACTTGTGACCCAGGA 3312
Qy      1322 Ser11eLeuAsnGlyThrArgVal11eGlySerAspLeuLeuGlnGlySerSerValThrTrp 1341
      3313 AATATATGATGATGAGACAGATTTGAGACACTTCAAGCTTGCTCCACATCACTCACTAC 3372
Db      1342 TyrCyHis11eGly11eTrpGluVal11eGlnGlyThrSerThrLeuSerCys11eLeuGlnPro 1361
      3373 CAGTGTGACTCTGGCTATTAAGATTTTGAACCCCTCATTCATCACTGCTGATGGGGCT 3432
Qy      1362 AspGlyLeuProVal11ePheAsnProArgProValCysThrAlaProCysGlyGln 1381
      3433 GATGGAGAAACCTTCCTGGAGCAAGTGTCCCTCTCTGCAATGCTCCCTGTGAGAGCCAG 3492
Db      1382 TyrVal11eGlySerAspGlyValValLeuSerProAsnTrpProGlnAsnTrpThrSerGly 1401
      3493 TACAGCGGATCAGAGGGATGATTTATACCAAACTTACCCCATATATACCACTGAT 3552
Qy      1402 Gln11eCysLeuTrpPheValThrValProGlyAspTrpValValPheGlyGlnPheAla 1421
      3553 CAATATGCTCTATTCATCACTCAAGGTAACAAAGAAATGTGTGCTTGGAGACATTTGCC 3612

```

```

Qy      1422 PhePheHisThrAlaLeuAsnAspValValGluValHisAspGlyHisSerGlnHisSer 1441
      3613 TATTTCAGACAGCCCTGATGATTTGGCAGATTTTGAATGAGAACCATGACAGAGCC 3672
Db      1442 ArgLeuLeuSerSerLeuSerGlySerHis11eGlnGlyLeuSerLeuProLeuAlaThrSer 1461
      3673 AGACTTCTAGCTCACTCTCGGGGTCTCACTGAGGAGAAACATGCTCTGCTGATGCTCA 3732
Qy      1462 AsnGlnValLeu11eLeuPheSerAla11eGlyLeuAlaProAlaArgGlyPheHisPhe 1481
      3733 AATCAATCTGCTCCGATTCACTGCAAGAGAGCGGTGCTTCCGCGGCTTCACTTC 3792
Db      1482 ValTrpGlnAlaValProArgThrSerAlaThrGlnCysSerSerValProGluProArg 1501
      3793 GTGTATCAAGCTGTTCCTCTGACAGTACACCCCAATGACGCTCTGCTCCAGGCCAGA 3852
Qy      1502 TyrGlyLeuArgLeuGlySerAspPheSerVal11eValAla11eValArgPheGlyCysAsn 1521
      3853 TACGAGAGAGATTTGTTCTGAGTTTCTGCGGCTCATGCTCATGCTGCAATGCAATGCA 3912
Db      1522 SerGlyTyrAlaLeuGlnGlySerProGlu11eGlnCyLeuProValProGlyAlaLeu 1541
      3913 CCGGATACCTGCTTCAAGGTTCCAGCGGCTCCAGCTGCACTGCACTGCGCCAAAGCCTTG 3972
Qy      1542 AlaGlnTrpAsnValSerAlaProThrCysValValProGlyGlyValAsnLeuThrGlu 1561
      3973 GACAGTGGAGACACATCCCACTGCTGTGTGATCACTGACAGTGAATTTCACTCA 4032
Db      1562 ArgArgGlyThr11eLeuSerProGlyPheProGluProGlyLeuAsnSerLeuAsnCyS 1581
      4033 CGAAGGATGACATCTGTCCTGCCCCGCTACCTGAGCCATACGAAACAACTTGAATGCT 4092
Qy      1582 ValTrpVal11eValValProGluGlyAlaGly11eGln11eGlnValValSerPheVal 1601
      4093 AATGAGAAATCATGATTAACGAGGCTCGGGAATTCAGATCAAGTATGATGATTTGCC 4152
Db      1602 ThrGluGlnAsnTrpAspSerLeuGlnVal11eAspGlyVal11eAspAsnThrValThrMet 1621
      4153 ACGAGAGCAATCTGGAATCTCTTGAATGATCCAGATGTGGAGATGTGACGCCAGCA 4212
Qy      1622 LeuGlySerPheSerGlyThrThrValProAlaLeuAsnSerThrSerAsnGlnLeu 1641
      4213 CTGGAGACCTTCTCAGGACACAGTACCGGCACTGCTGAGACATCTTCCAACTCACTC 4272
Db      1642 TyrLeuHisPheTrpSerAsp11eSerValSerAlaAlaGlyPheHisLeuGlnTrpYrs 1661
      4273 TACCTGCAATTCAGTGTGACATTAAGTGTGACAGCTGCTGTTTCCACCTGGAATACAAA 4332
Qy      1662 ThrVal11eGlyLeuSerSerCySProGluProAlaValProSerAsnGlyVal11eTrpGly 1681
      4333 ACTGATGCTCTTCTGCTGATGCAAGCAAGACAGCCCTCCCAAGCAACAGCATATAAATGGGA 4392
Db      1682 GluArgTrpLeuVal11eAspValValSerPheGlnCysGluProGly11eTrpAlaLeuGln 1701
      4393 GATCGGTATCAAGTGAACAGAGTGTCTCTCTTCAAGTGGAGCCGGGTACACCTCGAG 4452
Qy      1702 GlyHisAlaHis11eSerCysMetProGlyThrValArgArgTrpAsnTrpProProPro 1721
      4453 GCGCTTCCCAATTTCTGATGCAAGGAGACGTTGCGGTGGAACATATCCCTCTCCC 4512
Db      1722 LeuCy11eAlaGlnCysGlyGlyThrVal11eGlnLeuGlyVal11eLeuSerPro 1741
      4513 CTGTGATTTGCAACTGTGGAGGAGACCTAGACACTTGGGTGTGATCTCGAGCCCC 4572
Qy      1742 GlyPheProGlyLeuTrpProSerAsnMetAspCysSerTrpVal11eAlaLeuProVal 1761
      4573 GGTTCCTCAAGTTCTTACCCCAACTTATGACTGCACTGAGAGATCTCATTAACCATC 4632
Db      1762 GlyPheGlyAlaHis11eGlnPheLeuAsnPheSerThrGluProAsnHisAspTrp11e 1781
      4633 GGTATGTGCAATATTCAGTTTCTGAATTTTTCACGAACTATATCATGACTTCCCTT 4692

```

```
Qy 1782 Glu1IleArgAsnGlyProTyrGluThrSerArgMetSerGlyArgPheSerGlySerGlu 1801
Db 4693 GAAATTCAAAATGGAACCTTACACACAGCCCATGATGGACATTTAGGGGACGAT 4752
Qy 1802 LeuProSerSerLeuLeuSerThrSerHisGluThrValTyrPheHisSerAspHis 1821
Db 4753 CTCCCGGGGGCCCTGCTGAGCACAACGATGAACCCCTCATCTTTATATAGACAT 4812
Qy 1822 SerIleAsnArgProGlyPheLeuLeuGluTyrGlnIleValTyrGluLeuGlnGluPro 1841
Db 4813 TCGGAAACCGGCAAGATTTAACTTGCTTACCAAGCCTTAGAATTAACAACCTGCCA 4872
Qy 1842 AspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrAsnValGlyGlnSer 1861
Db 4873 GATCACCCCATTTTCAGAAATGGGTACATGATCACTGGATTCAGGTGAGGGGCAATCA 4932
Qy 1862 ValThrPheGluCysLeuProGlyTyrGlnLeuThrGlnHisProValLeuThrCysGln 1881
Db 4933 GTATCTTTCGAGTGTATCTGGGGTACATCTTAATAGGCCATCTGCTCATCTTGTCAG 4992
Qy 1882 HisGlyThrAsnArgAsnTyrAspHisProLeuProGlyCysGlyValProCysGlyGly 1901
Db 4993 CATGGGATTCACAACAACCTGGAACCTACCTTTCACAGATGTATGCCCTTGGGGTAC 5052
Qy 1902 AsnIleThrSerSerAsnGlyThrValTyrSerProGlyPheProSerProTyrSerSer 1921
Db 5053 AACGTAACTTCTCGAAGCGACACATCTACTCCCTGCTTCTGATGAGTATCCGATC 5112
Qy 1922 SerGlnAspCysValTyrLeuIleThrValProIleGlyHisGlyValArgLeuAsnLeu 1941
Db 5113 CTGAAGACCTGATTTGGCTCATCAGCGGTGCTCCAGGCGACGGAGTTTACATCACTTC 5172
Qy 1942 SerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleTyrAspGlyProGlnGln 1961
Db 5173 AACCTGTATCAGACGGAAGCTGTCAACGATTAATGCTGTGGAGCGGCTCCGATCAG 5232
Qy 1962 ThrAlaProArgLeuGlyValPheThrArgSerMetAlaGlySerThrValGlnSerSer 1981
Db 5233 AACTCACCCCGAGCTGGAGTTTTCAGTGGCAACAGCCCTCGAAACGGCGTATAGCTCC 5292
Qy 1982 SerAsnGlnValLeuLeuLysPheHisArgAspAlaIleThrGlyGlyIlePheAlaIle 2001
Db 5293 ACCAACCAAGCTCGCTCAAGTTCACAGCAGCATTTTCAAAATGAGGGCTTCTTGTCCTC 5352
Qy 2002 AlaPheSerAlaTyrProLeuThrLysCysProProProThrIleLeuProAsnAlaGlu 2021
Db 5353 AATTTCACGCAATTCAGCTCAGAAAGTCAACCTCCCGACGGGTTCCACAGCGAGAA 5412
Qy 2022 ValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrArgCysLeuPro 2041
Db 5413 ATGCTTACTGAGATGATGATTTGAGATAGAGATTTTGGAGTACCAAGTCCACCC 5472
Qy 2042 GlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGlnPhe 2061
Db 5473 GGGATACACCTTGGTGGGAGCCGACATTCGACTTGGCAAGCTCAAGTCCCAAGTTCAGTT 5532
Qy 2062 GluGlyProProProIleCysGlyValHisCysProThrAsnGluLeuLeuThrAspSer 2081
Db 5533 GAGGGTCTCTCCCAACATGTGAAGCACAATGCCCAGCAAAATGAAGTCCGAGCTGATCA 5592
Qy 2082 ThrGlyValIleLeuSerGlnSerTyrProGlySerTyrProGlnPheGlnThrCysSer 2101
Db 5593 TCGGAGATCATTCAGTCCAGGGATCCGGGTAAATTTTAACTCCCAAGACTTGCTCT 5652
Qy 2102 TyrLeuValArgValGluProAspTyrAsnIleSerLeuThrValGluTyrPheLeuSer 2121
Db 5653 TGGAGTATTAAGTGAACCAAACTACAACTTACATCTTGTGGACACATTTCAAAAT 5712
Qy 2122 GluLysGlnTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeuLeu 2141
Db 5713 GAAAGCAGTGTGTGACACTGGAAGTGTGTGATGTTCTTCTGGGCAAAAGTCCCTGCTA 5772
Qy 2142 LysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerSerAsnSerVal 2161
Db 5773 GTAGTCTTAAGTGGGATCATCTAGCAACATCAATTTTACAAAGCAGAGTAACTAGTCA 5832
Qy 2162 TyrLeuArgTyrPheSerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyrSer 2181
Db 5833 TATCTCCGCTGGTCCATCGACATGACCAATGCCAAGAAAGATTAACGATTCGCTATGCA 5892
Qy 2182 AlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlnThrSer 2201
Db 5893 GCACCTTACTGAGTGTGACCCACCCCTGAGATGGGGTATTTCTTAAACGAGACTGCA 5952
Qy 2202 ThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGlyHis 2221
Db 5953 GAGCGGTGGAGCAAAAGCATTTATTTTGGACGCTGGATACCGAATGTCGGGCAC 6012
Qy 2222 SerMetAlaIleCysThrArgHisProGlnGlyTyrHisLeuTyrSerGluAlaIlePro 2241
Db 6013 AGCAATCAACCTGTACACAAACCCACTTGGCATGTACAGTGGGACTCCCTCACGCCA 6072
Qy 2242 LeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysAsnGlyMetValPheGly 2261
Db 6073 CTGTGCGAGCTGTCTGTGTGATCCAGAAATCCCAAGAAACGGTTCATTTACCGGG 6132
Qy 2262 LysGluTyrThrValGlyThrLysAlaValTyrSerCysSerGlnGlyTyrHisLeuGln 2281
Db 6133 AACGATTCATTTGGACAGTAAAGTGTCTATGAATGTCATAGAGGCTTCACAGCTTGA 6192
Qy 2282 AlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuTyrSerAsnArgAsnVal 2301
Db 6193 TCCAGCAGACAGCAACACAGCCGTGTGTCAAGAAAGATGGCTGTGGAGTAAACAAGGGG 6252
Qy 2302 ProProGlnCysValProValThrCysProAspValSerSerIleSerValGluHisGly 2321
Db 6253 CCGCCACGCTAGCCCGGTGCTGTGCCCAAGATTAACCTCAGCTTCAGAAACAGTGC 6312
Qy 2322 ArgTyrArgLeuIlePheGluThrGlnTyrGlnPheGlnIleGlnMetLeuIleCys 2341
Db 6313 ATCTGAGAGCTGTTCAGATCTCTGAATGAGTACGCTCTCAAGTATTTGAGCTGC 6372
Qy 2342 AspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTyr 2361
Db 6373 AGTCTGTGTACTTACCTTAGAAGCTGGAGCTGCTCGCGGTCCAGGCCAATGGGACGTGG 6432
Qy 2362 SerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIlePro 2381
Db 6433 AACATAGAAATAGAGAGCCCAAGCTGTGAGTATCTGCTGTGGAACCTTCTCTTCC 6492
Qy 2382 ProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSerCys 2401
Db 6493 CCAATGCAACAAGATTTGAAAGCTTACAGATTTATGGGGCCACAGCTATATTTAGCTGC 6552
Qy 2402 AsnSerGlyTyrThrLeuValGlySerArgValArgGlyCysMetAlaAsnGlyLeuTyr 2421
Db 6553 AACACCGGTACACGCTTGTGGGCTCATGTGTGAGAGTGTGGCAAAATGGGCTGTGG 6612
Qy 2422 SerGlySerGlyValArgCysLeuAlaGlyHisCysGlyThrProGluProIleValAsn 2441
Db 6613 AGCGGACGCAAACTCATGTCTGGCGGACGTGCGCTGCCAGACCCGAGTTGTGAAC 6672
Qy 2442 GlyHisIleAsnGlyLeuAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAla 2461
Db 6673 GGTCACTTATGTGAGATGCTTCAAGTACAGACACGGGTGTTTACAGTGCATCTCT 6732
Qy 2462 GlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisIleTyrSerGly 2481
Db 6733 GGTTCGGGCTTGGGAACTTCCGTGAGATATGCTCGCAAAACCAAGTGTGTGGA 6792
Qy 2482 LysThrProPheCysValProIleThrCysGlyHisIleProGlyAsnProValAsnGlyLeu 2501
Db 6793 CAACCGCTCTGTGTGCTCCCATACATGTGGTACCTGTGAAACCTTGCCACGGAATTC 6852
Qy 2502 ThrGlnLysAsnGlnPheAsnLeuAsnArgValValLysPheValCysAsnProGlyTyr 2521
Db 2502 ThrGlnLysAsnGlnPheAsnLeuAsnArgValValLysPheValCysAsnProGlyTyr 2521
```

Db 6853 ACTAATGGCAGTAGCTTCACTGAAATGATGTCGTAATTCACCTGCACACGGGCTAT 6912
 QY 2522 MetAlaGluGlyValAlaArgSerGlnCysLeuAlaSerGlyGlnTyrPheAspMetLeu 2541
 Db 6913 TTGCTGCAGGGGCTGCTCTCGAGCCAGTGTGGAGCAACGGCCAGTGAAGAGCCCTG 6972
 QY 2542 ProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluAsnSerValArgGln 2561
 Db 6973 CCCAGCTGTGAGGGGTACACTGTTCTGATCCAGGCTTTGGGAAAATGCCATTCTGTCAC 7032
 QY 2552 ValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValSerTyrArgCysAsn 2581
 Db 7033 GGGCAACAACACTCCCTGAGAGATTTCAGATGAGTAAGATGATCTGTCATTCATTCAG 7092
 QY 2582 HisGlyPheTyrLeuLeuGlyThrProValLeuSerCysGlnGlyAspGlyThrTyrAsp 2601
 Db 7093 AAGGATTTTACTTCTGGGATCTTCAGCCTTGACCTGTATGGCAAAATGGCTTATGGAC 7152
 QY 2602 ArgProArgProGlnCysLeuLeuValSerCysGlyHisProGlySerProProHisSer 2621
 Db 7153 CGATCCCTGCCCAAGTGTGGTATATCGTGTGAGAACCCAGGGGCTCCCTGCCAAGCC 7212
 QY 2622 GlnMetSerGlyAspSerTyrThrValGlyAlaValAlaArgTyrSerCysIleGlyLys 2641
 Db 7213 GTCTCTCACTGAGAGCTGTTTACCTATGGCGCCCTGTCACATACTCTGACAGAGGGAGC 7272
 QY 2642 ArgThrLeuValGlyAsnSerThrArgMetCysGlyLeuAspGlyHisTyrPheArgLys 2661
 Db 7273 GAGAGCTCATAGCAACGACACGAGAGTGTCCAGAGAAAGCTCATCTGAGCCGGAGCA 7332
 QY 2662 LeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGlyIleProAlaHis 2681
 Db 7333 CTGCCCACTGCACAGGAAATTAATCTGATCTGTGTGATCCGGGAGCCCAAGACAT 7392
 QY 2682 GlyIleLeuLeuGlyAspSerPheAspProGlyThrValMetArgPheSerCysGlyAla 2701
 Db 7393 GGGCTCTGGCTGTGATGACTTAAACAACAAGAGCTTCTCCGCTTCTCTGTGAATG 7452
 QY 2702 GlyHisValLeuArgGlySerSerGlyArgThrCysGlnAlaAsnGlySerTyrPheArg 2721
 Db 7453 GGGCAGCAGCTGAGGGGCTCCCTGAAACGACGCTGTTCATGGTCAATGGTCAATGGTCA 7512
 QY 2722 SerGlnProGlyCysGlyValIleSerCysGlyAsnProGlyThrProSerAsnAlaArg 2741
 Db 7513 CTGACAGCGGTGTGTGAGCGCTGCTCTGTGGCAACCTGTGGCACACCAACGGAATG 7572
 QY 2742 ValValPheSerAspGlyLeuValPheSerSerIleValTyrGlnCysArgGlyLys 2761
 Db 7573 ATTGTCAGTAGTATGATGATCTTCTGCTCCAGCTCGGTCACTATGCTGTGGAAAGGC 7632
 QY 2762 TyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyThrTyrThrGlySer 2781
 Db 7633 TACAAAGACCTCAGGGCTCATGACACGGCATTTGCACGCAATGGAGCTCGACAGCACT 7692
 QY 2782 AspProGlyCysLeuValIleAsnCysGlyAspProGlyIleProAlaAsnGlyLeuArg 2801
 Db 7693 GCTCCGAGCTGCAATTAATGTTGTGGGATCCAGGACACACTGAATGGCAATGCCATCCAG 7752
 QY 2802 LeuGlyAsnAspPheArgTyrAsnLysThrValThrTyrGlnCysValProGlyIleTyrMet 2821
 Db 7753 TTTGGGACGCACTTCACTTCAACAAGACTGTGAGCTATCACTGATGATCCAGGCTATGTC 7812
 QY 2822 MetGlnSerHisArgValSerValLeuSerCysThrLysAspArgThrTyrAsnGlyThr 2841
 Db 7813 ATGGAGACGATCATCCGCTATTCGCTGTACCAAGAGCGGAGCGGGAATCCGAGC 7872
 QY 2842 LysProValCysLeuValAlaLeuMetCysLysProProProLeuIleProAsnGlyLysVal 2861
 Db 7873 AAACCTGTCTGCAAAACCGGTGTGTCTTCAGCCGCCCGGTGAGATGGAACAGTGTG 7932
 QY 2862 ValGlySerAspPheMetTyrGlySerSerValThrTyrAlaCysLeuGlnGlyTyrGln 2881
 Db 7933 GAGGGAAGTGAATTTCCGCTGGGCTCCACATTAAGTATACGCTGCATGAGCGGTTACAG 7992

QY 2882 LeuSerLeuProAlaValPheThrCysGlnGlyAsnGlySerTyrThrGlyLeuPro 2901
 Db 7993 CTCTCTCACTCCGGCCATCTCTCTGTGAAGTCCGGGGGTGTGGAAAGAGATCCCC 8052
 QY 2902 GlnCysPheProValPheCysGlyAspProGlyValProSerArgGlyArgArgGluAsp 2921
 Db 8053 CAGTGTCTCCCTGTGTGTGGGAGACCTGTGCATCCCGCAGAAAGGGGCACTTGTAGTGG 8112
 QY 2922 ArgGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuValLeuVal 2941
 Db 8113 AAAAGTTTCACTATTAAGTCCGAGACTCTTCTCCAGTGCATTCATCTTACTCTGTG 8172
 QY 2942 GlySerProArgArgPheCysGlnSerAspGlyThrTyrPheSerGlyThrGlnProSerCys 2961
 Db 8173 GGATCTCCAGAAAGTGTGCCCAACCTGACAGCGCATGTGAGCGGCATCAACCCACTTC 8232
 QY 2962 IleAspProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyIleGlnAsn 2981
 Db 8233 ATTTGATCTGCTCATTAACACTGCCCCAGACCTGTGTACGCCACACTTGGAAATACAGAT 8292
 QY 2982 AsnSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlnLysGlyTyrLeu 3001
 Db 8293 AGCTCCAGAGGCTATGAGGTGGAAAGCAGGTTTTTTCAGGTGCAGAAAGCTTACAT 8352
 QY 3002 LeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTyrPheSerGlyThrProPro 3021
 Db 8353 ATTCAGAGTTCACACACCTGACACCTGCTTCCATTTAATGATGAGTGGATACAGACC 8412
 QY 3022 AspCysValProHisHisCysArgGlnProGlyIleThrProThrHisAlaAsnValGlyAla 3041
 Db 8413 GAATTAATACCTCAAGCTGCGACAGCAGCAAAACCCGGCAGACGGGATGTGAGACC 8472
 QY 3042 LeuAspLeuProSerMetGlyTyrThrLeuIle-ThrProAlaArgArgAlaSerProSe 3061
 Db 8473 ATCGATCTTCTTACTTGTGGCTACACCTTATGTATACCTGCATCCAGGCTTTTTCCTC 8532
 QY 3061 PArgValAlaProSerThrAlaProAlaArgArgMetAlaIleGlyGlnAlaSerArg 3080
 Db 8533 GCAAGGAGATCTGAGCAGCAAGATGTAACAGACATGAATGACAGAAAGTCCG 8590

RESULT 13
 AAD33318
 ID AAD33318 standard; cDNA; 10673 BP.
 AC AAD33318;
 DT 01-JUN-2002 (first entry)
 DE Human C3b/C4b complement receptor like cDNA #1.
 KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
 KW transplant rejection; autoimmune disease; ischemic condition; nocturnal;
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
 KW infertility; vasodilator; obesity; cardiac; gene; ss.
 OS Homo sapiens.
 PH Location/Qualifiers
 FT CDS 334..9543
 FT /tag= a
 FT /product= "Human C3b/C4b complement receptor like protein
 #1"
 FT /transl_except= (pos:2416..2418, aa:Xaa)
 FT /note= "This translational exception occurs while
 FT decoding the alternative version of human C3b/C4b
 FT complement receptor like protein #1 (AAE20900)"
 PN W0200210199-A2.
 XX

PD 07-FEB-2002.
 XX
 XX 24-JUL-2001; 2001WO-US023222.
 XX
 XX 02-AUG-2000; 2000US-0222504P.
 PR 28-NOV-2000; 2000US-00728787.
 XX
 PA (AMGE-) AMGEN INC.
 PI Welcher AA, Elliott GS;
 XX
 XX WPI; 2002-303934/34.
 DR P-PSDB; AAE20787, AAE20900.
 XX
 PT Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
 PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
 PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
 PT sclerosis.
 XX
 PS Claim 1; Fig 1; 251pp; English.
 XX
 CC The invention relates to a nucleic acid encoding a novel C3b/C4b
 CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
 CC polypeptide and nucleic acid molecules may be used to treat, prevent,
 CC ameliorate, diagnose and/or detect diseases such as immune system
 CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
 CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's
 CC disease), ischemic conditions, metabolic disorders (e.g. obesity and
 CC diabetes) and infertility. The invention is useful in gene therapy. The
 CC present sequence is human C3b/C4b complement receptor like cDNA
 XX
 XX Sequence 10673 BP; 2690 A; 2711 C; 2574 G; 2691 T; 0 U; 7 Other;
 SQ
 Alignment Scores:
 Pred. No.: 0 Length: 10673
 Score: 10737.00 Matches: 1864
 Percent Similarity: 78.37% Conservative: 433
 Best Local Similarity: 63.60% Mismatches: 567
 Query Match: 63.21% Indels: 67
 DB: Gaps: 7
 US-10-016-248-2 (1-3104) x AAD33318 (1-10673)
 QY 163 ValTrpIleIleThrAlaLeuAsnProSerLyValIleLysLeuAlaPheGluGluPhe 182
 Db 202 GTTGGCTGTTTAAAGCAGAACCCCTCTTCAGTG-----GAAGTC 246
 QY 183 Asp-----LeuGluArgGlyTrpAspThrLeuThrValGlyAspGlyGln 198
 Db 247 GATGTAATGTTGGCTTAAAGCTTAAAGCTTGAC----- 279
 QY 199 AspGlyAspGlnLysThrValLeuTyxMetSerGlnAsnAlaCysSerAspSerProHis 218
 Db 280 -----ATTTCCTTTCTTGACG 297
 QY 219 ThrProGlySerArgGlyLeuProGlu-----SerMetSerGlyAspIleTrpArgGln 235
 Db 298 CTCACGGGATCAGAGTCTGACCTCATGTGAGCAGTCAACAGCATGTGCTCAT 357
 QY 236 LysTrpThrValLeuGluIleCysArgAspIleSerSerSerAspAlaArgSerGlySer 255
 Db 358 -----CTGACGTCGATGATGATGAC----- 375
 QY 256 ValArgLysSerProLysThrSerAsnAlaValGluLeuValAlaProGly----- 272
 Db 376 -----ATTGCTCACTCGGTTAAAGCT 399
 QY 273 -----ThrGluIleGluGlnGlySerCysGlyAspProGlyIleProAlaTyxGlyArg 290
 Db 400 GTTACCAAGAAATTGAAAGGAGGAGGTGTGGGATCCTGGAATCCCGCTATGGAGAG 459

QY 291 ArgGluLysSerArgPheHisIleGlyAspThrLeuLysPheGluCysGlnProAlaPhe 310
 Db 460 CGAAGCGGCGAGCAATTTCCCTCCATGGAGATACATCACTTGAATCCCGCGGCTTT 519
 QY 311 GluLeuValGlyGlnLysAlaIleThrCysGlnLysAsnAngIntPSSerAlaLys 330
 Db 520 GAGCTGTGGGGAGAGATTAATCACTGTACAGAGAACATCAGTGTGGCAAGAG 579
 QY 331 ProGlyCysValPheSerCysPhePheAsnPheThrSerProSerGlyValAlaLeuSer 350
 Db 580 CCCAGCTGTATATTTTCATGTTCTTCACTTAAAGCAGCATCATCTGGATTTATCTGTCA 639
 QY 351 ProAsnTrpProGluAspTrpGlyAsnHisIleCysValTrpLeuIleLeuAlaArg 370
 Db 640 CCAATATATCAAGAGAAATATGGAGACAAATGAACTGTCTGTGATTAATCTGGAG 699
 QY 371 ProGluSerArgIleHisIleLeuAlaPheAsnAspIleAspValGluProGlnPheAspPhe 390
 Db 700 CCAAGAAATCGAATTCATCTTAATCTTAAATGATTTGATGTGAGCTCAATTTGACTTT 759
 QY 391 LeuValIleLysAspGlyAlaThrAlaGluAlaProValLeuGlyThrPheSerGlyAsn 410
 Db 760 CTCGCGGTCAAGATGATGATGATTCGATTAATCTGCTGGTACTTTTCTGGCAAT 819
 QY 411 GluLeuProSerSerIleThrSerSerGlyHisValAlaArgLeuGluPheGlnThrAsp 430
 Db 820 GAAATGCTTTCCAGCTGCGCAGAGCATGGGCAATTAATCTGTTGAATTTCACTCTGAC 879
 QY 431 HisSerThrGlyLysArgGlyPheAsnIleThrPheThrThrPheArgHisAsnGluCys 450
 Db 880 CATTCCACTACGTGCAAGAGGTTCAACATCATCTTACACCAATTGTGTGCAATGATGTC 939
 QY 451 ProAspProGlyValProValAsnGlyLysArgPheGlyAspSerLeuGlnLeuGlySer 470
 Db 940 CATGATCTCGCATTCCTTAATAACGACGATTTGTGACAGTTTCACTCGGAGC 999
 QY 471 SerIleSerPheLeuCysAspGluGlyPheLeuGlyThrGlnGlySerGluThrIleThr 490
 Db 1000 TCGGTTCTTTCCACTGTGATGATGATGCTTTGTGACAGCCAGAGATCCGATTCATACC 1059
 QY 491 CysValLeuLysGluGlySerValIleTrpAsnSerAlaValLeuArgCysGluAlaPro 510
 Db 1060 TGCACTACGCAAGCGGGAAGTGTGTGAGCTCCACGTCGCCGCTGGAAGCTCCA 1119
 QY 511 CysGlyGlyHisIleuThrSerProSerGlyThrIleLeuSerProGlyTrpProGlyPhe 530
 Db 1120 TGTGTGTGACATCTGACAGCGTCCAGCGAGTCAATTTGCTCTGTGATGCCAGGATAT 1179
 QY 531 TyrIleAspAlaLeuSerCysAlaTrpValIleGluAlaGlnProGlyTyxProIleLys 550
 Db 1180 TATAGGATTCCTTACATTGTGATGATGATATGAAACCAAAACAGGCCACTCTATCAAA 1239
 QY 551 IleThrPheAspArgPheLysThrGluValAlaAsnTyxAspThrLeuGluValArgAspGly 570
 Db 1240 ATAACCTTTACAGATTTCAGACAGAGTCAATTAATGACCTTGAGAGTCAAGATGAGG 1299
 QY 571 ArgThrTyxSerAlaProLeuIleGlyValTyxHisGlyThrGlnValProGlnPheLeu 590
 Db 1300 CCAAGCAGTTCTGTCCTCCACATGATCGGAGTACCAAGCCAGAGCAACCCAGTTCTC 1359
 QY 591 IleSerThrSerAsnTyxLeuTyxLeuLeuPheSerThrAspLysSerHisSerAspIle 610
 Db 1360 ATCAGACCGGGAATTCATGATCTGCTATTCACCTACCAACAGCCGCTCAGATC 1419
 QY 611 GlyPheGlnLeuArgTyxGluThrIleThrLeuGlnSerAspHisCysLeuAspProGly 630
 Db 1420 GGCTTCTCATCACTATGAGAGTGAAGTGAACCTTGAGTCCCTGCGACCCGGCG 1479
 QY 631 IleProValAsnGlyGlnArgHisGlyAsnAspPheTyxValGlyAlaLeuValThrPhe 650
 Db 1480 ATCCCTGTGAACGACATCCCAAGGTGAGACTTTGGCATCGTCCACAGTGAACATTTT 1539
 QY 651 SerCysAspSerGlyTyxThrLeuSerAspGlyGluProLeuGluCysGluProAsnPhe 670

```
Db 1540 AGCTGTGACCGGGGTATACACTAAGTACACAGACCGCCCTGCTGTGAGAGAACCCAC 1599
Qy 671 GlnTPSerAArgAlaLeuProSerCysGluAlaLeuCysGlyGlyPheIleGlnGlySer 690
Db 1600 CAGTGGAAACCAAGCGCTTGCACAGCTGCACCGCTCATATGTGAGAGGCTCATCAAGGAG 1659
Qy 691 SerGlyThrIleLeuSerProGlyPheProAspPheTyrProAsnAsnLeuAsnCysThr 710
Db 1660 AGTGAACAGTCTTTCTCTCGGGTTTCCAGATTTTATCCAACTCTCTTAACCYGACG 1719
Qy 711 TrpIleIleGlnThrSerHisGlyValPhePheThrPheHisIleu 730
Db 1720 TGGACATTTGAAGTGTCTCATGGGAAGAGATTCAAAATCTTTCACCTTTCATCTT 1779
Qy 731 GluSerGlyHisAspTyrLeuLeuIleThrGluAsnGlySerPheThrGlnProLeuArg 750
Db 1780 GAGAGTTCACAGACTATTATCTGATCAACAGAGATGGAAGTTTTCGAGCCGCTTGGC 1839
Qy 751 GlnLeuThrGlySerAArgLeuProAlaProIleSerAlaGlyLeuTyrGlyAsnPheThr 770
Db 1840 AGGCTCACCGGGGTGGGTGCTTCATACCATCAAGCAGCGCTTGGAACTTCTACT 1899
Qy 771 AlaGlnValAArgPheIleSerAspPheSerMetSerTyrGlnGlyPheAsnIleThrPhe 790
Db 1900 GCCACGCTTCGGTTTATATACAGACTTCTCAATTTCGTAAGAGGCTTCAATATCAGATT 1959
Qy 791 SerGlnTyrAspLeuGluProCysGlnGluProGluValProAlaTyrSerIleArgLys 810
Db 1960 TCAGAAATATGACTCGAGCCAGATGATGATCTGGAATCTCTGCTTCCACCCAGAAATT 2019
Qy 811 GlyLeuGlnPheGlyValGlyAspThrLeuThrPheSerCysPheProGlyTyrArgLeu 830
Db 2020 GGTTTTCATTTGGTGTGGAGACTCTCTGACGTTTTCGCTTCTCGGGAATCTGTTA 2079
Qy 831 GluGlyThrAlaArgIleThrCysLeuGlyValArgArgLeuThrSerSerProLeu 850
Db 2080 GAAGGTCCRCACAGCTTACCTGCTGGGTGGGGCGCGCTGTGTGAGTGCACCTCTG 2139
Qy 851 ProArgCysValAlaGluCysGlyAsnSerValThrGlyThrGlnGlyThrLeuLeuSer 870
Db 2140 CCAAGGTGTGCGCCAGATGTGAGCAAGTGTCAAAAGAAATAGAAAGACATTACTGTCT 2199
Qy 871 ProAsnPheProValAsnTyrAsnAsnAsnIleGlyCysIleTyrSerIleGlnThrGln 890
Db 2200 CCAAAATTTTCCATCCATATGTGATTAATACATGAGTGTATCTATAAATAGAAACAGA 2259
Qy 891 ProGlyLysGlyIleGlnLeuLysAlaArgAlaPheGlnLeuSerGlyLysAspValLeu 910
Db 2260 GCCGGCAAGGCGATCCACTTACAAACAGAGCTTCAGCTGTGTAAGAGATACTCTTA 2319
Qy 911 LysValTyrAspGlyAsnAsnAsnSerAlaArgLeuGlnGlyValPheSerHisSerGln 930
Db 2320 AAGGTATATGATGAAAAGACAGTCTCTACGTCACCTGGGACGTTTCACTAAAAATGAA 2379
Qy 931 MetMetGlyValThrLeuAsnSerThrSerSerSerLeuThrLeuAspPheIleThrAsp 950
Db 2380 CTCTGAGGCTGATCTCTAAACAGACATCCATCACTGTGCTGAGATCTTCAACACCAAT 2439
Qy 951 AlaGlnAsnThrSerLysGlyPheGlnLeuHisPheSerSerPheGlnLeuIleLysCys 970
Db 2440 GGAATCGACACCAACAGATTTCACCTCATACAGATTTGATGTGTAATAATGT 2499
Qy 971 GluAspProGlyThrProLysPheGlyTyrLysValHisAspGlnGlyHisPheAlaGly 990
Db 2500 GAGGATCCGGGCGATCCCTTAACCTAGGCTATAGGATCCGATGAGGACCTTTACCGAC 2559
Qy 991 SerSerValSerPheSerCysAspProGlyTyrSerLeuArgGlySerGlnGluLeu 1010
Db 2560 ACTGTAATTTCTGTACAGTTCACCCGGGGTACGGCATGTGAGGACGAACACCTTGACC 2619
Qy 1011 CysLeuSerGlyGluArgArgThrTyrAspArgProLeuProThrCysValAlaGlnCys 1030
Db 2620 TGTTTAGTGGACACGAGAGTGTGGACAAACCTACTCTGTGCATAGCGGAATGT 2679
Qy 1031 GlyGlyThrValArgGlyGluValSerGlyGlnValLeuSerProGlyTyrProAlaPro 1050
Db 2680 GGTGTCAATCATCATTACACCCACCATCAGACCAAAATTTCTCCCTGTGCTATCCAGTCCG 2739
Qy 1051 TyrGlnHisAsnLeuAsnCysIleThrThrIleGlnAlaGluAlaGlyCysThrIleGly 1070
Db 2740 TATGACAAACCTCCACTGCACCTGGATTATAGAGGCGAGACCAGAAAGACATTAC 2799
Qy 1071 LeuHisPheLeuValPheAspThrGlnGluValHisAspValLeuAspGlyLeuThrAspGly 1090
Db 2800 CTTCATTTATTTGTTTGCACACGAGATGCTCACAGATCTCTCAAGATCTCAAGGTCTGGACGG 2859
Qy 1091 ProValGluSerGlyValLeuLeuLysGlyLeuSerGlyProAlaLeuProLysAspLeu 1110
Db 2860 CCGGTGACAGTGCATCTGCTGAAGAGTGAAGTGGCTCGGCCCTTCGGAGAGACATC 2919
Qy 1111 HisSerThrPheAsnSerValValLeuGlnPheSerThrAspPhePheThrSerLysGln 1130
Db 2920 CACAGACCTTCACTCACTCACCTGCAGATTCACAGCACTTCTTATCAGCAAGTCT 2979
Qy 1131 GlyPheAlaIleGlnPheSerValSerThrAlaThrSerCysAsnAspProGlyIlePro 1150
Db 2980 GGCTTCTCATCAGTCTTCCACTCATTCAGGCCACTGTACGATTCAGATATGCC 3039
Qy 1151 GlnAsnGlySerArgSerGlyAspSerTrpGluAlaGlyAspSerThrValPheGlnCys 1170
Db 3040 CAAAAATGGACCCGCTATAGAGACAGCAGAGGCTGAGACACCGTCACATTCCAGTGT 3099
Qy 1171 AspProGlyTyrAlaLeuGlnGlySerAlaGlnIleSerCysValLysIleGluAsnArg 1190
Db 3100 GACCTGTGCTACCTCCAAAGACACCAAAATCACTGTGTGCACCTGAATTAACGG 3159
Qy 1191 PhePheTrpGlnProSerProProThrCysIleAlaProCysGlyGlyAspLeuThrGly 1210
Db 3160 TTCTTTGGCAACGACCCCTCTCATAGATGCTGTTGTGAGGGAATCTGACGGGC 3219
Qy 1211 ProSerGlyValIleLeuSerProAsnTyrProGluProTyrProProGlyLysGlnCys 1230
Db 3220 CCAGAGGTGTATTTGTGCACCAACCAACAGCGGTTCCTCCGGGAGAAATGT 3279
Qy 1231 AspTrpLysValThrValSerProAspTyrValIleAlaLeuValPheAsnIlePheAsn 1250
Db 3280 GACTGGAAGTAAGTAAGTAACCCGAGCTTTGTCTGCTGATTAATCAAAAGTTTCAAC 3339
Qy 1251 LeuGlnProGlyTyrAspPheLeuHisIleTyrAspGlyArgAspSerLeuSerProLeu 1270
Db 3340 ATGAGGCCACCTATGACTTCTCTACATCTATGAAAGGGAATTCACAGCCCCCTC 3399
Qy 1271 IleGlySerPheTyrGlySerGlnLeuProGlyArgIleGlnSerSerSerAsnSerLeu 1290
Db 3400 ATTGGAGTTTCCAGGGGTCTCAGAGCCCAAGAAAGTAAGATACCGGAACAGCTCG 3459
Qy 1291 PheLeuAlaPheArgSerAspAlaSerValSerAsnAlaGlyPheValIleAspTyrThr 1310
Db 3460 TTTCTGGGATTTGGAGATGATGCTCGGTGGCTTTCAGGGTTCGCCATTTGAATTAA 3519
Qy 1311 GluAsnProAlaGluSerCysPheAspProGlySerIleLysAsnGlyThrArgValGly 1330
Db 3520 GAGAAACCAACGGAAGCTTTGTTGACCCAGAAATTAATGATGAGCAACAGATTGGA 3579
Qy 1331 SerAspLeuLysLeuGlySerSerValThrTyrTyrCysHisGlyGlyTyrGlnValGlu 1350
Db 3580 ACAACATTCACAGCTTGGCTCCACATATCACTTACAGTGTACTCTGGCTTAAAGATTCTT 3639
Qy 1351 GlyThrSerThrLeuSerCysIleLeuGlyProAspGlyLysProValTyrAsnAsnPro 1370
Db 3640 GACCCCTCATCCATCACCTGTGATTTGGGCGTGAAGGAACCTCTCGGACCAAGTG 3699
Qy 1371 ArgProValCysThrAlaProCysGlyGlyGlnTyrValGlySerAspGlyValLeu 1390
Db 3700 CTGCCCTCTGCATGCTCTCTGTGAGGCCAGTACACGGGATCAGAAAGGATGTTT 3759
```

QY 1391 SerProAsnTyrProGlnAsnTyrThrSerGlyGlnIleCysLeuTyrPheValThrVal 1410
 DB 3760 TCACCAATCAATCCCAATATTTACACAGCTGGTCAAAATATGCTTATTCATCCAGGTA 3819
 QY 1411 ProLysAspTyrValValPheGlyGlnPheAlaPhePheHisThrAlaLeuAsnAspVal 1430
 DB 3820 CCAAGAAATTCGTGGTCTTTGGACAGTTTGGCTATTTCCAGACAGCCCTGAATGATTTG 3879
 QY 1431 ValGluValHisAspGlyHisSerGlnHisSerArgLeuLeuSerSerLeuSerGlySer 1450
 DB 3880 GCAGAAATTAATTTGATGGAACCATGACAGGCAACATTTCTCAGTCTCCTCGGGGCT 3939
 QY 1451 HisThrGlyGlnSerLeuProLeuAlaThrSerAsnGlnValIleLeuIlePheSerAla 1470
 DB 3940 CACTCAGGGGAAACCTTGGCTTGGCTACGTCAAAATTCATATTCGTCTCGATCCAGTGA 3999
 QY 1471 LysGlyLeuAlaProAlaArgGlyPheHisPheValTyrGlnAlaValProArgThrSer 1490
 DB 4000 AAGAGCGGTGCTGCTGGCCGGGCTTCCATTCGTATCAAGCTGTTCTCGTACCACT 4059
 QY 1491 AlaThrGlnCysSerSerValProGlnProArgTyrGlyValGlyLeuGlySerAspPhe 1510
 DB 4060 GACACCCCATGCTGCTGTCCCGAGCCAGATACGAGAGAGATTGTTCTGAGTTT 4119
 QY 1511 SerValGlyAlaIleValAlaArgPheGlyCysAsnSerGlyTyrAlaLeuGlnGlySerPro 1530
 DB 4120 TCTCCCGCTCCATGCTCCGATTCGATGACAAACCGGGAATACCTGCTTCAAGGTTCCAG 4179
 QY 1531 GluIleGlyCysLeuProValProGlyAlaLeuAlaGlnTyrAsnValSerAlaProThr 1550
 DB 4180 GCGCTCCACTGCTCCAGTCCGCTCCCAAGCCCTTGGCAGCTGGAACGACATCCCAAC 4239
 QY 1551 CysValValProCysGlyGlyAsnLeuThrGluArgArgGlyThrIleLeuSerProGly 1570
 DB 4240 TGTGTGATACCTGCTGAGTGGCAATTTCACTCAACGAGAGGTACATCTCTCCCGGC 4299
 QY 1571 PheProGlnProTyrLeuAsnSerLeuAsnCysValTyrPheValIleValProGlnGly 1590
 DB 4300 TACCTTGAACCATACGAAACAATCTGAATGTAATGAAATCATATGATACGAGGCT 4359
 QY 1591 AlaGlyIleGlnIleGlnValValSerPheValThrGlnAsnTyrAspSerLeuGlu 1610
 DB 4360 TCGGGAATTCAGATCCAAAGTATCATGTTTGCACGAGCAGAACTGGACTCTCCTTACG 4419
 QY 1611 ValPheAspGlyAlaAspAsnThrValThrMetLeuGlySerPheSerGlyThrThrVal 1630
 DB 4420 ATCCACGATGTGGGATGTGACCGCACCCAGACTGGGAGACTTCTCAGCACACAGTA 4479
 QY 1631 ProAlaLeuLeuAsnSerThrSerAsnGlnLeuTyrLeuHisPheTyrSerAspIleSer 1650
 DB 4480 CCGGCACTGCTGAACAAGTATCTCAACCACTTACCTGCAATTTCCAGTTCGACATTAAT 4539
 QY 1651 ValSerAlaAlaGlyPheHisLeuGlnTyrTyrSerThrValGlyLeuSerSerCysProGlu 1670
 DB 4540 GTGGCAGTGTGTTTCCACTCGAATACAAACTGTAGTCTTTCGTCGATGCCAAGAA 4599
 QY 1671 ProAlaValProSerAsnGlyValIleThrGlyGluArgTyrLeuValAsnAspValVal 1690
 DB 4600 CCACCCCTCCCAACACAGCATCAAAATCGAGATCGGTACATGATGAAAGCATGCTCTC 4659
 QY 1691 SerPheGlnCysGlnProGlyTyrAlaLeuGlnGlnHisAlaHisIleSerCysMetPro 1710
 DB 4660 TCTTTCACATGTCAGCCCGGGTACACCTGCAAGGCGCTTCCACATTTCTGTATGCA 4719
 QY 1711 GlyThrValAlaArgTyrPheAsnTyrProProProLeuCysIleAlaGlnCysGlyThr 1730
 DB 4720 GGGACCGTTCGCTGGAACTATTCCTGCTCTCCCTGTGCATTTGCACTGTGGAGGAGC 4779
 QY 1731 ValGluGluMetGluGlyValIleLeuSerProGlyPheProGlyAsnTyrProSerAsn 1750
 DB 4780 CTGAGCACTTGGTGGTGTGATCTCTGAGCCCGGCTTCCAGAGTTCTTACCCCAACAAC 4839

QY 1751 MetAspCysSerTyrPheIleAlaLeuProValGlyPheGlyAlaHisIleGlnPheLeu 1770
 DB 4840 TTAACATGCACTGAGAGATCTCATTTACCAATCGGCTATGTGACATATTCAGTTTCTG 4899
 QY 1771 AsnPheSerThrGlnProAsnHisAspTyrIleGlnIleArgAsnGlyProTyrGluThr 1790
 DB 4900 AATTTTCTACCGAAGCAATCATGACTTCTCTTGAATTCAAATGAGACCTTACACACC 4959
 QY 1791 SerArgMetMetGlyAlaArgPheSerGlySerGlnLeuProSerSerLeuLeuSerThrSer 1810
 DB 4960 AGCCCATGATGTGACAAATTTAGCGCAAGATTCCTCCCGGCTCTGTAGACAAACG 5019
 QY 1811 HisGluThrThrValTyrPheHisSerAspHisSerGlnAsnArgProGlyPheLeuLeu 1830
 DB 5020 CATGAACCCCTCATCCACTTTTATATGACCATTTCCGAAACCGGCAAGATTTAACTT 5079
 QY 1831 GluTyrGlnAlaTyrGluLeuGlnGluCysProAspProGluProPheAlaAsnGlyIle 1850
 DB 5080 GCTTACCAAGCTATGATTAACAGAACCTGTCAGATCCACCCCATTCAGAAATGGGTAC 5139
 QY 1851 ValArgGlyAlaGlyTyrAsnValGlyGlnSerValThrPheGluCysLeuProGlyTyr 1870
 DB 5140 ATGATCAACTCGGATTCAGGTGGGCAATCAGATATCTTTCAGTGTATTCCTGGGTAC 5199
 QY 1871 GlnLeuThrGlyHisProValLeuThrCysGlnHisGlyThrAsnArgAsnTyrAspHis 1890
 DB 5200 ATTCTATATAGCCATCTGCTGCTCATCTTTCAGATGGATCAACAGAACTGGAATAC 5259
 QY 1891 ProLeuProLysCysGluValProCysGlyGlyAsnIleThrSerSerAsnGlyThrVal 1910
 DB 5260 CTTTTCAGATGTGATGAGCCCTGTGGGTAACAAGTAACTTCTCAGAACCGCACCATC 5319
 QY 1911 TyrSerProGlyPheProSerProTyrSerSerSerGlnAspCysValTyrPheIleThr 1930
 DB 5320 TACTCCCTCGCTTCTGATGAGTATCCATCTCGTAAGACATGCTCATTTGGCTCATACG 5379
 QY 1931 ValProIleGlyHisGlyValArgLeuAsnLeuSerLeuGlnThrGlnProSerGly 1950
 DB 5380 GTGCTTCAGGCAAGGAGTTTACATCACTTCACTGTTACAGAGGAAAGCTGTCAAC 5439
 QY 1951 AspPheIleThrIleTyrPheGlyProGlnGlnThrAlaProArgLeuGlyValPheThr 1970
 DB 5440 GATTAACATGCTGTTTGGACGGTCCCGATCAGAACTCACCCCAAGCTGGAGTTTCACT 5499
 QY 1971 ArgSerMetAlaValValThrValGlnSerSerSerAsnGlnValLeuLeuLysPheHis 1990
 DB 5500 GGCACACACGCTCGAAACGGCGTATAGCTCCACCAACAAAGTCTGCTCAAGTTCCAC 5559
 QY 1991 ArgAspAlaAlaThrGlyGlyIlePheAlaIleAlaPheSerAlaTyrProLeuThrLys 2010
 DB 5560 AGCGACTTTTCAATGAGAGGCTTCTTGTCTCAATTTCCACGCAATTTCAAGCTCAAGAA 5619
 QY 2011 CysProProProThrIleLeuProAsnAlaGluValValThrGluAsnGluGluPheAsn 2030
 DB 5620 TGTCAACTCCCCAGCGGTTCCACACAGCAGAAATGTTACTGAGATGATGATTTGAG 5679
 QY 2031 IleGlyAspIleValAlaGlyTyrArgCysLeuProGlyPheThrLeuValGlyAsnGluIle 2050
 DB 5680 ATAGAGATATTTTGAAGTACAGTCCAGTCCCGGAGTACCTTGTGGGAGCCGACAT 5739
 QY 2051 LeuThrCysValLeuGlyThrTyrTyrLeuGlnPheGlnGlyProProPheIleCysGluVal 2070
 DB 5740 CTGACTTGAAGCTCACTTCCAGTTGACATTTGAGGTTCTCTCCCAACATGGAAGCA 5799
 QY 2071 HisCysProThrAsnGlnLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTyr 2090
 DB 5800 CAATGCCCAAGATAGAGCCCGGACGATCATTCGGAATCATTTCTCATCCAGGTTAT 5859
 QY 2091 ProGlySerTyrProGlnPheGlnThrCysSerTyrPheValArgValGluProAspTyr 2110
 DB 5860 CCGGTATATTAATTTAACTCCAGACTTGTCTTGGAGTATTAAGTGAACCAAACTAC 5919
 QY 2111 AsnIleSerLeuThrValGluTyrPheLeuSerGluLysGlnTyrAspGluPheGluIle 2130

```
Db 5920 AACATTACCATCTTTGTGGACACATTTCAAGTGAAGAGCTTTGATGACCTGGAAGTG 5979
Qy 2131 PheAspGlyProSerGlyGlnSerProLeuLeuYalaLeuSerGlyYAsnTyrSerAla 2150
Db 5980 TTTGATGGTTCTTCTGGGCAAGCTCTCTAGTACTTTAAGTGGGAATCATATCAAGAA 6039
Qy 2151 ProLeuLeuValThrSerSerSerAsnSerValTyrLeuArgTyrPsrSerSerAsnAla 2170
Db 6040 CAATCAATTTTACAGACAGAGATATCATATATCTCCGCTGCTCATGACATGACCAAGCC 6099
Qy 2171 TyrAsnArgGlySerGlyPheLeuValArgTyrSerAlaProTyrCysSerLeuProArgAla 2190
Db 6100 ACCAGTAAGAAAGATTCAGATTCAGCTATGACACCTTACTGACATTTGACCAACCC 6159
Qy 2191 ProLeuHisGlyPheLeuLeuGlnThrSerThrGlnProGlyGlySerLeuHisPhe 2210
Db 6160 CTGAAGATGGGGTATTTCTTAACAGACTGACAGACGGGTGGACCAAGATGCAATTAT 6219
Qy 2211 GlyCysAsnAlaGlyTyrArgLeuValGlyHisSerMetAlaLeuCysThrArgHisPro 2230
Db 6220 TTTTGCAAGCCTGGATATCCGAATGTCGGCCACGCAATGCACTGTATGACAAACCA 6279
Qy 2231 GlnGlyTyrHisLeuTyrPsrSerGlnAlaLeuProLeuCysGlnAlaLeuSerCysGlyLeu 2250
Db 6280 CTGGCATGTACACAGTGGGACTCCCTCACCCACTGCGACGGCTGTCTGTGGATC 6339
Qy 2251 ProGlnAlaProLeuYAsnGlyMetValPheGlyLeuGlyTyrThrValGlyThrValAla 2270
Db 6340 CCAGAAATCCCGAGAAACGGTTCTTTACCGGGAACAGATTCACCTTTGGACAGTAAAGTG 6399
Qy 2271 ValTyrSerCysSerGlnGlyTyrHisLeuGlnAlaGlyValAlaGlnAlaThrAlaGlnCys 2290
Db 6400 GTCTATGAATGTATGAGGGCTTCAAGCTTGAATCCAGCAGCAAGAAAGCCGTGTGT 6459
Qy 2291 LeuAspThrGlyLeuTyrPsrSerAsnValAlaProProGlnCysValProValThrCys 2310
Db 6460 CAAAGAAATGGCTGTGGATTAACAAAGGGAGCCGCCACTGTAAAGCCGCTGCCCTGC 6519
Qy 2311 ProAspAlaSerSerLeuSerValGlnHisGlyArgTyrArgLeuLeuPheGlnThrGln 2330
Db 6520 CCAGACATTTAAGCTCAGCTCTCAGAAACATGTATCTGAGGCTGTTCAGAGATCTTGG 6579
Qy 2331 TyrGlnPheGlnAlaGlnLeuMetLeuLeuCysAspProGlyTyrTyrTyrThrGlnGln 2350
Db 6580 AATGATGACGGTCTCAAGATATGTGAGCTGACGCTGCTGTTACTTAAAGGCTGCG 6639
Qy 2351 ArgValIleLeuArgCysGlnAlaAsnGlyLysTyrPsrSerLeuGlyAspSerThrProThrCys 2370
Db 6640 AGGCTCTCGGGTCCAGGCAATGGAAGCTGGAACATAGAGATGAGAGGCCAAGCTGT 6699
Qy 2371 ArgIleLeuSerCysGlyGlnLeuProIleProProAsnGlyHisArgIleGlyThrLeu 2390
Db 6700 CGAGTTATCTCGTGGAGAGCTTTCTTCCCCCAATGAGCAACAGATTGGAAGCTTG 6759
Qy 2391 SerValTyrGlyAlaThrAlaLeuPheSerCysAsnSerGlyTyrThrLeuValGlySer 2410
Db 6760 ACGATTATAGGGGCCACAGCTATATTACGTGCAACACCGGCTACACGCTTGGGGTCT 6819
Qy 2411 ArgValArgGlyCysMetAlaAsnGlyLeuTyrPsrSerGlySerGlnValArgCysLeuAla 2430
Db 6820 CATGTACAGAGAGCTTTGGCAATGGGCTCTGGAGGGGCAAGCACTGAGTGTGGGT 6879
Qy 2431 GlyHisCysGlyThrProGlnProIleValAsnGlyHisIleAsnGlyGlnLysTyrSer 2450
Db 6880 GGCACATGCGGTTCCCAAGCCGATTTGAAACGGTCACTAAGTGAAGATGGCTTCACT 6939
Qy 2451 TyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIleGlyMetSerVal 2470
Db 6940 TACAGAGACAGGTGGTTTACAGAGTCAATCTGTGTTCCGGCTTGGGAACTTCCGTG 6999
Qy 2471 ArgIleCysGlnGlnAspHisIleTyrPsrSerGlyLysThrProPheCysValProIleThr 2490

Db 7000 AGGATATGCCCTGCAGACCAAGATGCTTGCAGAAACCCCTGTCTGTCTCCCATCACA 7059
Qy 2491 CysGlyHisProGlyYAsnProValAsnGlyLeuThrGlnGlyYAsnGlnPheAsnLeuAsn 2510
Db 7060 TGTGTACACCTCGAAGAACCTCCGACAGATTCACATTAAGGCAAGTCAACTCAAGT 7119
Qy 2511 AspValIleValPheValCysAsnProGlyTyrMetAlaGlnGlyAlaAlaArgSerGln 2530
Db 7120 GATGTGTGAATTTTCACTCGCAACACGGGCTATTTGCTGCGAGGGGCTGTGAGCCAG 7179
Qy 2531 CysLeuAlaSerGlyGlnTyrPsrSerAspMetLeuProThrCysArgIleLeuAsnCysThr 2550
Db 7180 TGTGAGACAGACGGCCAGTGGAGTACCTCTGCCACCTGTGCTGAGTGAACGTGTT 7239
Qy 2551 AspProGlyHisGlnGlnAsnSerValArgGlnValHisAlaSerGlyProHisArgPhe 2570
Db 7240 GATCCAGGCTTTGTGGAATAATCCATTCGTATCGTCAAGGCAACAGAACTCCCTAGAGTTT 7299
Qy 2571 SerPheGlyThrThrValSerTyrArgCysAsnHisGlyPheTyrLeuLeuGlyThrPro 2590
Db 7300 GAGTATGAAATGAGATCTCTTACCATTCAGAAAGGAATTTTACTTGTGGATCTTCA 7359
Qy 2591 ValLeuSerCysGlnGlyYAspGlyTyrThrAspArgProArgProGlnCysLeuLeuVal 2610
Db 7360 GCCTTGACTGTATGGAATAATGCTTATGGACCGATCCCTGCCAAGTGTGGCTTAT 7419
Qy 2611 SerCysGlyHisProGlySerProProHisSerGlnMetSerGlyYAspSerTyrThrVal 2630
Db 7420 TCGTGTGACACCCAGGGGCTCTGCGCAACCCGCTCTCATGAGAGACTGTATCTAT 7479
Qy 2631 GlyAlaValAlaArgTyrSerCysIleGlyLysArgThrLeuValGlyYAsnSerThrArg 2650
Db 7480 GCGCGCTGTGTGACTACTCTCTGACAGAGGACGAGAACCTCATATGGGCAACACGAGA 7539
Qy 2651 MetCysGlyLeuAspGlyHisIleTyrThrGlySerLeuProHisCysSerGlyYThrSerVal 2670
Db 7540 GTGTCCAGAAAGACATGACATGAGGGGCGGACCTGCCCATGACAGAAATATCTCT 7599
Qy 2671 GlyValAlaCysGlyYAspProGlyTyrLeuProHisIleGlyTyrLeuGlnGlyYAspSerPheAsp 2690
Db 7600 GGATTCGTGTGTATCCCGGAGACCCAGACATGGTGTCTCGCTTGTGATGACTTTAAG 7659
Qy 2691 ProGlyThrValMetArgPheSerCysGlnAlaGlyHisValLeuArgGlySerSerGln 2710
Db 7660 ACAAGAGTCTTCTCCGCTTCTCTGTGAATGGGGACCAAGCTGAGGGCTCTCCCTGAA 7719
Qy 2711 ArgThrCysGlnAlaAsnGlySerTyrPsrSerGlySerGlnProGlnCysGlyValIleSer 2730
Db 7720 CGCAGTGTGTTCATTAATGGGTATGATGAGGACTGACAGACTGACGGGTGTGAGCCGTGCC 7779
Qy 2731 CysGlyYAsnProGlyYThrProSerAsnAlaArgValAlaPheSerArgGlyLeuValPhe 2750
Db 7780 TGTGCAACCTGCGACACCCACCAAGGAATGATTTCTGATGATGATGATGATGATGATGAT 7839
Qy 2751 SerSerSerLeuValTyrGlnCysArgGlnGlyTyrTyrAlaThrGlyLeuLeuSerArg 2770
Db 7840 TCCAGCTCGGACATCTATGCTGCTGGGAAGGCTTACAAAGACTGAGGCTCATACACGG 7899
Qy 2771 HisCysSerValAsnGlyTyrThrGlySerAspProGlnCysLeuValIleAsnCys 2790
Db 7900 CATTGACAGGCAATGGGACTGAGACAGGACTGCTCCGACATGCACAATTAATTAAGTTGT 7959
Qy 2791 GlyYAspProGlyIleProAlaAsnGlyLeuArgLeuGlyYAsnAspPheArgTyrAsnLys 2810
Db 7960 GGGGATCCAGGACACATAGCAAAATGGCATCAAGTTGGGACCACTTCACTCAACAG 8019
Qy 2811 ThrValThrTyrGlnCysValProGlyTyrMetMetGlnSerHisArgValSerValLeu 2830
Db 8020 ACTGTAGCTATAGTGTAACTCAGGCTATGCTATGGAACAGATCATCCGACATAT 8079
Qy 2831 SerCysThrTyrAspArgThrTyrPsrSerGlyThrLysProValCysValAlaLeuMetCys 2850
Db 8080 CGCTGTACAAAGACGGCAGGTGGAATCCGAGCAAACTGTCTGCAAAAGCCGTGCTGTGT 8139
```


QY 2851 LysProProLeuLeuProLeuValValGlySerAspPheMetTrpGlySer 2870
 DB 8140 CCTACGCGCGCGCGCGAGAAATGAGAGAGAGAGATTCCTCGCGAGCTCC 8199
 QY 2871 SerValThrTyrAlaCysLeuGluGlyTyrGlnLeuSerLeuProAlaValPheThrCys 2890
 DB 8200 AGCATTAAGTTACAGCTGATGATGAGAGGTTTACAGCTCTTCACTCCGATCTCTCTGT 8259
 QY 2891 GluGluAsnGlySerTrpThrGlyGluLeuProGlnCysPheProValPheCysGlyAsp 2910
 DB 8260 GAAAGTCGCGGCGGCTGAGAAAGAGAGATCCCGAGTGTCTCCGTGTCTCGCGAGAC 8319
 QY 2911 ProGlyValProSerAspArgLysArgGluAspArgGlyPheSerTyrArgSerSerVal 2930
 DB 8320 CTTGGCAATCCCGGAGAAAGGCGACTTATGAGGAAAGTTTACCTTAAAGTCCGAAAGTC 8379
 QY 2931 SerPheSerCysHisProProLeuValLeuValGlySerProArgArgPheCysGlnSer 2950
 DB 8380 TTCTTCAGTGCATTAATCTCCATTTATTAATCTCGTGGATCTCCAGAGAGTCTGCCAAGCT 8439
 QY 2951 AspGlyThrTyrPheSerGlyThrGlnProSerCysIleAspProThrLeuThrCysAla 2970
 DB 8440 GACGCGACGTGAGCGGATCAACCCACCTGATGATCTGCTCATTAACACCTGCCCA 8499
 QY 2971 AspProGlyValProGlnPheGlyTyrIleAsnAsnSerGlnGlyTyrGlnValGlySer 2990
 DB 8500 GACCTGGTACGACCACTTTGAAATACAGATAGCTTCAGAGGCTATGAGGTTGAGAGC 8559
 QY 2991 ThrValLeuPheArgCysGlnLysGlyTyrLeuLeuGlnGlySerThrThrArgThrCys 3010
 DB 8560 ACGGTTTTTTTTCAGGTGAGAAAGGCTACCATATTCAGGTTCCAGACCTCGACCTGC 8619
 QY 3011 LeuProAsnLeuThrTyrPheSerGlyThrProProAspCysValProHisCysArgGln 3030
 DB 8620 CTTCCCAATTTAAATGATGAGTGGATGAGATCAGACCAATGTATATCACTGCTGACAGAG 8679
 QY 3031 ProGluThrProThrHisAlaAsnValGlyValLeuAspLeuProSerMetGlyTyrThr 3050
 DB 8680 CCAGAAACCCCGGACGCGGAGTGGATGAGAGCCATGATCTTCTTCTGAGTACACC 8739
 QY 3051 LeuIle-ThrProAlaArgArgAlaSerProSerArgValAlaProSerThrAlaProAl 3070
 DB 8740 TTACTGTACACCTGCCATCCAGGCTTTTCTCCAGAGGAGATCTGACACAGAAACATGT 8799
 QY 3070 AArgArgMetAlaAlaGlyGlnAlaSerArg 3080
 DB 8800 AAGCAGACATGAATGAGAGAAAGTCCG 8830
 RESULT 14
 AAD33319
 ID AAD33319 standard; cDNA; 12525 BP.
 XX AAD33319;
 AC
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Rat C3b/C4b complement receptor like cDNA.
 XX
 KW Rat; C3b/C4b complement receptor-like molecule; immune system disorder;
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
 KW transplant rejection; autoimmune disease; ischaemic condition; nocturnal;
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
 KW infertility; vasodilator; obesity; cardiac; gene; ss.
 XX
 OS Rattus rattus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..9288
 FT /*tag= a
 FT /product= "Rat C3b/C4b complement receptor like protein"

FT /note= "CDS does not include start codon"
 FT /partial
 XX
 FN MO200210199-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 24-JUL-2001; 2001MO-US023232.
 XX
 PR 02-AUG-2000; 2000US-0222504P.
 PR 28-NOV-2000; 2000US-00728787.
 PA (AMGE-) AMGEN INC.
 XX
 PI Welcher AA, Elliott GS;
 XX
 DR WPI; 2002-303934/34.
 DR P-PSDB; AAE20788.
 XX
 PT Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
 PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
 PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
 PT sclerosis.
 PS
 PS Claim 1; Fig 3; 251pp; English.
 XX
 CC The invention relates to a nucleic acid encoding a novel C3b/C4b
 CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
 CC polypeptide and nucleic acid molecules may be used to treat, prevent,
 CC ameliorate, diagnose and/or detect diseases such as immune system
 CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
 CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's
 CC disease), ischaemic conditions, metabolic disorders (e.g. obesity and
 CC diabetes) and infertility. The invention is useful in gene therapy. The
 CC present sequence is rat C3b/C4b complement receptor like cDNA
 XX
 SQ Sequence 12525 BP; 3136 A; 3208 C; 3002 G; 3106 T; 0 U; 73 Other;
 SQ
 Alignment Scores:
 Pred. No.: 0 Length: 12525
 Score: 10720.00 Matches: 1843
 Percent Similarity: 79.52% Conservative: 436
 Best Local Similarity: 64.31% Mismatches: 543
 Query Match: 63.11% Indels: 44
 DB: 6 Gaps: 5
 US-10-016-248-2 (1-3104) x AAD33319 (1-12525)
 QY 195 AspGlyGlyGlnAspGlyAspGlnValLeuTyrMetSerGlnAsnAlaCysSer 214
 DB 1 GATGCCGGAAGGCGGAGACACACATCCGCTCTTACGCTG----- 42
 QY 215 AspSerProHisThrProGlySerArgLysProGlu-----SerMetSerGlyAsp 231
 DB 43 -----CTTACAGGCTCGAGTCTCTGACTCTCACTGATCGTACATGACATCAG 90
 QY 232 IleThrArgGlnLysTrpThrValLeuGluIleCysArgAspIleSerSerSerAspAla 251
 DB 91 ATGCGGCTCCAC-----CTGCAGTCAAGACGACAGC 120
 QY 252 ArgSerGlySerValArgLysSerProLysThrSerAsnAlaValGluLeuValAlaPro 271
 DB 121 -----ATTGGTCCCA 132
 QY 272 Gly-----ThrGluIleGlnGlySerCysGlyAspProGlyIlePro 286
 DB 133 GGAATTTAAAGCTGTATCAAGAAATGAGAGAGAGCTGCGGAGCTCGGATCCCA 192
 QY 287 AlaTyrGlyArgArgGlnGlySerArgPheHisIleGlyAspThrLeuLysPheGluCys 306
 DB 193 GCCTACGGGAAGGAGATGAGACAGAGCTTTTCCACGCGGAGACAGCTCACTTGAAGTGC 252

QY 307 GlnProAlaPheGlnLeuValGlyGlnIleValIleThrCysGlnIleAsnAsnGlnIleTrp 326
 DB 253 CAGGCAAGCTTTGAGCTGTGAGAGAGAGATGATTTAGCTGCCAAGAAACCAACAGAGG 312
 QY 327 SerAlaValLeuProGlyCysValPheSerCysPheAsnPheThrSerProSerGly 346
 DB 313 TCCGGCAACAGCCAAAGCTGTGTGTTTCACTTTCTTCACTTCAAGCGCTCTCGG 372
 QY 347 ValValIleLeuSerProAsnThrProGlnAspTyrGlyAsnHisIleuHisCysValIlePleu 366
 DB 373 ATCATCTCTGTCCCAAACTATCTTAGAGAAATTTGGCAACACATGAATTTGTGTGTG 432
 QY 367 IleLeuAlaArgProGlnSerArgIleHisLeuAlaPheAsnAspIleAspValGlyPro 386
 DB 433 ATTATATCTGAGCCCGGAGCGGATTCACCTCATCTTCAATGATTTGAGTGTGAGCT 492
 QY 387 GlnPheAspPheLeuValIleLeuAspGlyAlaThrAlaGlnAlaProValIleuGlyThr 406
 DB 493 CAGTTTGACTTCTTCCGCTCAMAAGATGAGGATTTCTGACATCACAGTCTCTCGGACT 552
 QY 407 PheSerGlyAsnGlnLeuProSerSerIleThrSerSerGlyHisIleValAlaArgLeuGln 426
 DB 553 TTCTGTGCAATGAGAGTCTCTGACAGCTGCGCAGCTGAGACATAGTACGCTTGAG 612
 QY 427 PheGlnThrAspHisSerThrGlyValAspGlyPheAsnIleThrPheThrThrPheArg 446
 DB 613 TTTCAGTCCGATCACTTACACAGGAGGTTCAACATCATATACACACACTTTTGT 672
 QY 447 HisAsnGlnCysProAspProGlyValProValAsnGlyValAspPheGlyAspSerLeu 466
 DB 673 CAGAACAGAGTCTATGACCTTGGAGATCCCTGTGATGAGACGCGCTTTGAGACAGGTTT 732
 QY 467 GlnLeuGlySerSerIleSerPheLeuCysAspGlnGlyPheLeuGlyThrGlnGlySer 486
 DB 733 CTGCTGGAGAGTCTGTGCTCTTCCACTGATGATGCTTTGAGAGATCAGGCTTCT 792
 QY 487 GlnThrIleThrCysValLeuValGlnGlySerValValIleTrpAsnSerAlaValIleuArg 506
 DB 793 GAGCTATCATCATCATCTTGGCAAGATGAAACCTGTCTGAGACTCTATGTCCTCTCGC 852
 QY 507 CysGlnAlaProCysGlyGlyHisLeuThrSerProSerGlyThrIleLeuSerProGly 526
 DB 853 TGTAACTCTCTTGTGTGGGCACTGACAGCTTCTTGTGGGCTATATTAATCTTCAGGA 912
 QY 527 TrpProGlyPheTyrIleAspAlaLeuSerCysAlaTrpValIleGlnAlaGlnProGly 546
 DB 913 TGGCAGAGATATTAACAAGATCTTTAATTGCGAATGGGCTCATGACCAACAGAGA 972
 QY 547 TyrProIleLeuValIleThrPheAspArgPheLeuThrGlnValAsnTyrAspThrLeuGln 566
 DB 973 CATTCCTCATAAATATACATTTGACAGGTTCCAGACAGAGTCAATATATATCTCGGAA 1032
 QY 567 ValArgAspGlyArgThrTyrSerAlaProLeuIleGlyValTyrHisGlyThrGlnVal 586
 DB 1033 GTCCGGAGTGGCCAAACAGCTCATCCCACTGATTTGGGAGAGACATGGACCCAGGCT 1092
 QY 587 ProGlnPheLeuIleSerThrSerAsnTyrLeuTyrLeuLeuPheSerThrAspIleSer 606
 DB 1093 CCAAGTCTCTCATGACACAGGAACTCATGATGCTCTGTTTACCACTGACAGCAGC 1152
 QY 607 HisSerAspIleGlyPheGlnLeuArgTyrGlnThrIleThrLeuGlnSerAspHisCys 626
 DB 1153 CGGCGTAGTGTGGCTTCTCTCATCCCATGAGAGTGTACTTGAATCTGATCTCGT 1212
 QY 627 LeuAspProGlyIleProValAsnGlyGlnArgHisGlyAsnAspPheTyrValGlyAla 646
 DB 1213 CTGGACCCGCGGATCCCTGTAATAGTCTATCGGCACTGGCAATACCTTGTGATCATGACT 1272
 QY 647 LeuValThrPheSerCysAspSerGlyTyrThrIleuSerAspGlyGlnProLeuGlnCys 666
 DB 1273 ACAGTACCTTCACTGCTGATCCCTGGTACAGCTCATGATGAGATGCCCTCATCTGT 1332

QY 667 GlnProAsnPheGlnIleTrpSerArgAlaLeuProSerCysGlnAlaLeuCysGlyGlyPhe 686
 DB 1333 GAGAAAGAACCATCATGAGGAACACAGCCTTGCCAGCTGTGATGCTGTGTGAGGCTAC 1392
 QY 687 IleGlnGlySerSerGlyThrIleLeuSerProGlyPheProAspPheTyrProAsnAsn 706
 DB 1393 ATCATGAAAGAGTGGAGCTGTCTTTTACACAGATTTTCCAGACTTTTATATCAACTCT 1452
 QY 707 LeuAsnCysThrTrpIleIleGlnThrSerHisGlyIleGlyValPhePheThrPheHis 726
 DB 1453 CTGAATCTGATACAGACCATTTGAGTCTCTCATGGCAAGGAGTGCAGATGAATTTCCAC 1512
 QY 727 ThrPheHisLeuGlnSerGlyHisAspTyrLeuLeuIleThrGlnAsnGlySerPheThr 746
 DB 1513 ACCTTACACTTGAAGATCCACAGCATATTTGTGTATCAGACAGAGATGGAGTTTCTCA 1572
 QY 747 GlnProLeuArgGlnLeuThrGlySerThrArgLeuProAlaProIleSerAlaGlyLeuTyr 766
 DB 1573 GAGCCGATGCCAGGCTCATGGATCGCTCTGCTCACACCATTAAGCTGGCTTGT 1632
 QY 767 GlyAsnPheThrAlaGlnValArgPheIleSerAspPheSerMetSerTyrGlnGlyPhe 786
 DB 1633 GGAAACTTCACTGCGCAACTCAGATTCTCTGACTTCTCATCTCTCATATGAAGCTTC 1692
 QY 787 AsnIleThrPheSerGlyThrAspLeuGlnProCysGlnGlnProGlnValProAlaTyr 806
 DB 1693 AACATTTAGCTTTGCAAAATATGACCTTGAACCTGTGTATGACCTGTGAGATCCCTGCTTAC 1752
 QY 807 SerIleArgGlyGlyLeuGlnPheGlyValGlyAspThrLeuThrPheSerCysPhePro 826
 DB 1753 AGTCCAGAAATTTGGTTCAGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1812
 QY 827 GlyTyrArgLeuGlnGlyThrAlaArgIleThrCysLeuGlyGlyArgArgArgLeuTrp 846
 DB 1813 GGATACCGCTTATGAAAGTGCAACCAAGCTTACCTGTGTGTGTGTGTGTGTGTGTGTGT 1872
 QY 847 SerSerProLeuProArgCysValAlaGlnCysGlyAsnSerValIleThrGlyThrGlnGly 866
 DB 1873 AGTGCACTCTGCCAGAGT 1932
 QY 867 ThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsnHisGlyCysIleTyrSer 886
 DB 1933 ACATTACTCTCTCCAAATTTCCATCAATATATATATATATATATATATATATATATAA 1992
 QY 887 IleGlnThrGlnProGlyValGlyIleGlnLeuValAlaArgAlaPheGlnLeuSerGln 906
 DB 1993 ATGAAACAGAGCCGGAAGGGGATCATCTCAGACCCGGAACCTTCCAACTTTCGAA 2052
 QY 907 GlyAspValLeuValTyrAspGlyAsnAsnAsnSerAlaArgLeuLeuGlyValPhe 926
 DB 2053 GGAGACACTTAAAGTTTATGATGGAAGAGACAGCTCTCTGAGGTCACTGGAGTCTTC 2112
 QY 927 SerHisSerGlyMetMetGlyValThrLeuAsnSerThrSerSerIleuTrpLeuAsp 946
 DB 2113 ACAAGAAAGTAACTGAAGGGGCTGTGCTTAAACAGACCTTCCAAACCTGAGGCTGAG 2172
 QY 947 PheIleThrAspAlaGlnAsnThrSerTyrGlyPheGlnLeuHisPheSerSerPheGln 966
 DB 2173 TTCAACTCTAAGCGGTACAGTATCCGCCCAAGCTTTCAGCTTCACTTCAACCAATTTTGAC 2232
 QY 967 LeuIleCysGlnLeuAspProGlyThrProLysPheGlyTyrIleValHisAspGlnGly 986
 DB 2233 CTAGTAAATGATGAGATCCAGGATCCCTTAATAGGTGCTACAGATCCGAGATGATGCT 2292
 QY 987 HisPheAlaGlySerSerValSerPheSerCysAspProGlyTyrSerLeuArgIleSer 1006
 DB 2293 CACTTCAAGACACTGTGTCTTCTTACACTGACCAACCGGCTTACCAATGATGATGACAC 2352
 QY 1007 GlnGlnLeuLeuCysLeuSerGlyGlnArgArgThrTrpAspArgProLeuProThrCys 1026
 DB 2353 AGTACCTGACCTGTGATGGGAGCCGAAAGGTGTGGAAACAACTATGCTTCTCTGT 2412
 QY 1027 ValAlaGlnCysGlyGlyThrValArgIleGlnValSerGlyGlnValLeuSerProGly 1046

QY 1767 ILleGlnPheLeuAenPheSerThrGluProAenHisAspTyrIleGluIleArgAsnGly 1786
DB 4633 ATCCAAATTTCTGAATTTCTCACTGAAGCCAAACATAGTACTCGAGATCCAGAAAGGC 4692
QY 1787 ProTyrGluThrSerArgMetMetGlyArgPheSerGlySerGluLeuProSerSerLeu 1806
DB 4693 CCTTACCAAGTAGTCCAAATGATGGGACAGTTAGTGGCCCTGACCTGGCTGCCTCAG 4752
QY 1807 LeuSerThrSerHisGluThrThrValTyrPheHisSerAspHisSerGlnAsnArgPro 1826
DB 4753 CTGGACACCAACACATGAAACCCCTCATCCGCTTCTATGTGACCACTACACAGAACCCGCA 4812
QY 1827 GlyPheHisLeuGluTyrGlnAlaTyrGluIleuGlnGluCysProAspProGluProPhe 1846
DB 4813 GGATTTAACTCACTATCCAAAGCTTATGAGTTACAGAACTGCCGAGACCCCGCATTC 4872
QY 1847 AlaAsnGlyIleValAlaArgGlyAlaGlyTyrAsnValGlyGlnSerValThrPheGluCys 1866
DB 4873 CAGAAATGGGTTCAATGATCAACTCCGATTAACGGGTGGGACAGTCGATCTCATTTGAGTGC 4932
QY 1867 LeuProGlyTyrGlnLeuThrGlnHisProValLeuThrCysGlnHisGlyThrAsnArg 1886
DB 4933 TACCCGGGCTACATCTTGCTAGGACCCCTGTGCTACCTGCAGACATGGCAGCTGACAG 4992
QY 1887 AsnTyrAspHisProLeuProLysCysGluValProCysGlyGlyAsnIleThrSerSer 1906
DB 4993 AACTGGAACATACCTTTCCACAGGCTGAGCGCTCCCTGTGGATATATGAGATACAG 5052
QY 1907 AsnGlyThrValTyrSerProGlyPheProSerProTyrSerSerSerGlnAspCysVal 1926
DB 5053 AATGGCACCATTTATTCCTGGGTTCCACAGAGATCAATTCGAAAGGACCTCCG 5112
QY 1927 TyrLeuIleThrValProIleGlyHisGlyValArgLeuAsnLeuSerLeuGlnThr 1946
DB 5113 TGCTGTCTACTGCTCCTCCAGACATGAGTGTACATCACTTCACTTGCAGACT 5172
QY 1947 GluProSerGlyAspPheIleThrIleTyrAspGlyProGlnGlnThrAlaProArgLeu 1966
DB 5173 GAGGCTGTAAATGATCAATCGCTGTGTGGATGTCCTGACCAAGATCGCTCAGCTC 5232
QY 1967 GlyValPheThrArgSerMetAlaLysThrValGlnSerSerSerAsnGlnValLeu 1986
DB 5233 GGGGCTTCTAGTGAACACTGCTCCGAGACAGCATACACCTCCACCAACGAGCTTG 5292
QY 1987 LeuLysPheHisArgAspAlaAlaThrGlyGlyIlePheAlaIleAlaPheSerAlaTyr 2006
DB 5293 CTCAATTCACAGCGATTTCTCCATGAGGCTTCTTGTCTCAATTTTCATGCAATT 5352
QY 2007 ProLeuThrLysCysProProProThrIleLeuProAsnAlaGluValValThrGluAsn 2026
DB 5353 CAACCTAAGAGTGGCCGCTCCTCCAGTAGTCCGACAGCTGACCTGTCACAGAAAT 5412
QY 2027 GluGluPheAsnIleGlyAspIleValArgTyrArgCysLeuProGlyPheThrLeuVal 2046
DB 5413 GAAAGCTTTGAATTAAGGAGCTTGTAAAGTACCACTGCCATCCAGAGGTACACGCTTG 5472
QY 2047 GlyAsnGluIleLeuThrCysLysLeuGlyThrTyrLeuGlnPheGlnGlyLysProPro 2066
DB 5473 GGAAGTAGACACCTGATGACAGAGCTCAGCTCAAGCTATGTGTCAAAGGCTCCACCT 5532
QY 2067 IleCysGluValHisCysAspProThrAsnGluLeuLeuThrAspSerThrGlyValIleLeu 2086
DB 5533 ACCTGTGAAGACAAAGGCCAGCAATGAAAGTGCAGAAAGAGCTTCTGGGTGATTCTC 5592
QY 2087 SerGlnSerTyrProGlySerTyrProGlnPheGlnThrCysSerTyrLeuValArgVal 2106
DB 5593 AGTCTCGGTAACCCAGGACACTATTTAACTCCACAGACATGTGCTGAGTATTAAGG 5652
QY 2107 GluProAspTyrAsnIleSerLeuThrValGluTyrPheLeuSerGlyLysGlnTyrAsp 2126
DB 5653 GAGCCAAACTTAACTTAGGCTTGTGGACACCTTTCAAGAGTAAAGCAATTGAT 5712

QY 2127 GluPheGluIlePheAspGlyProSerGlyGlnSerProLeuLeuLysAlaLeuSerGly 2146
DB 5713 GCACCTGAAGATTTGATGATGTTCTTGGGCAAAAGCTTGTGTGATGCTTAAAGTGG 5772
QY 2147 AsnTyrSerAlaProLeuIleValThrSerSerSerAsnSerValTyrLeuArgTyrPsr 2166
DB 5773 AACCACTGAACAGTCCCAATTTTACACAGAAATACATATGATCTCCGCTGGCTGC 5832
QY 2167 SerAspHisAlaTyrAsnArgLysGlyPheLysIleArgTyrSerAlaProTyrCysSer 2186
DB 5833 ACAGATCATGACCAACCAAGAAAGATTCAGATTCGCTATGACGCTCTTACAGCAC 5892
QY 2187 LeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThrSerThrGlnProGlyGly 2206
DB 5893 CTACCTCTACACTCAAGAAATGGTGGCGTTTAAATTAACCGCAGGCGCCCTGGAGAC 5952
QY 2207 SerIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGlyHisSerMetAlaIleCys 2226
DB 5953 AAGGTGACGATATTCTGCAAGCTGGATATGCAATGATGGCCACAGACGCCACTGC 6012
QY 2227 ThrArgHisProGlnGlyTyrHisLeuTyrPheSerGluAlaIleProLeuCysGlnAlaLeu 2246
DB 6013 AGCGGAAACCCAGTGGGCTGACCAAGTGGGACTCGATGGACCCGCTTGGCAGGCTGG 6072
QY 2247 SerCysGlyLeuProGluAlaProLysAsnGlyMetValPheGlyLysGluTyrThrVal 2266
DB 6073 TCTGTGAAATTCACAGAGCTCCAGAAATGGCTCTTCAACAGCAATGATGATCACTTA 6132
QY 2267 GlyThrLysAlaValTyrSerCysSerGluGlyTyrHisLeuGlnAlaGlyValAlaGluAla 2286
DB 6133 GACAGTAAAGTACATTATGATATGTAAGAGCTTCAAGCTGGAAGCCAGTCAGACAAAC 6192
QY 2287 ThrAlaGluCysLeuAspThrGlyLeuThrPheSerAsnArgAsnValProProGlnCysVal 2306
DB 6193 ACTGTGTGTGCAAGAAAGATGGCTGTGGAGCAACAGAGAAAGCCACCAACGTCGCAA 6252
QY 2307 ProValThrCysProAspValSerSerIleSerValGluHisGlyArgTyrArgPheLeu 2326
DB 6253 CCGGTGCCCTGCCAGATGGAAGGCCAGCTGCACAGACACGCTCTGAGAGCTGGT 6312
QY 2327 PheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyrTyr 2346
DB 6313 TCGGATCATTAATGATATATGAGCTCAAGCTTCTCCACAGTGAAGTCTGGCTACTTC 6372
QY 2347 TyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTyrPheLeuGlyAspSer 2366
DB 6373 TTGAGGCTCAGAGGCTGTTCAGTGCACAGCCCAATGGGACCTGGAACAGAGAC 6432
QY 2367 ThrProThrCysArgLleIleSerCysGlyGluLeuProIleProProAsnGlyHisArg 2386
DB 6433 AGACCCAGATGTAAAGTCATCTCTGTGGAAGCTGTTCCTCCCAATGGTAAACAG 6492
QY 2387 IleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSerCysAsnSerGlyTyrThr 2406
DB 6493 ATAGGAGGCTCAATGATATGAGGACCAACCCCATCTTCACTCAATACCGGCTACACA 6552
QY 2407 LeuValGlySerArgValArgGluCysMetAlaAsnGlyLeuThrPheSerGlySerGluVal 2426
DB 6553 CTGTAGGCTCCCATGTCGGAGAGTCTGGCCAGATGCTCTGAGCGGATCTGAAACA 6612
QY 2427 ArgCysLeuAlaGlyHisCysGlyThrProGluProIleValAlaAsnGlyHisIleAsnGly 2446
DB 6613 AAGTGCCTGGGGCTCATTTGGCTCCAGAACCCCATTTGATGAGCCATATAGTGGC 6672
QY 2447 GluAsnTyrSerTyrArgGlySerValValTyrGlnCysAsnAlaGlyPheArgLeuIle 2466
DB 6673 GATGCTTCACTACAGAGACACAGTGTCTACCAAAAGCAACCTGGGTTTGCATCGTA 6732
QY 2467 GlyMetSerValArgGlyCysGlnGlnAspHisHisIleThrSerGlyLysThrProPheCys 2486
DB 6733 GGCAGCTGTGAGGATTTGCTGCAGACACCAAGTGTGGGGGACACCCCGCTTTCG 6792
QY 2487 ValProIleThrCysGlyHisProGlyAsnProValAsnGlyLeuThrGlnGlyAsnGln 2506

```

Db      6793 GTCCCATCATGTGGACACCTGGAAACCTGCGCATGCGCTCACCACGACGAG 6852
Qy      2507 PheanleuanaapvalVallypPheValCyasnProglyTyrmelaaglulValA 2526
Db      6853 TTCAACCTGATGACCTTGTGAATTTCACCTGCATCGGGCTACCTGCGAGGGTCC 6912
Qy      2527 AlaargserGlnCyaleuAlaserGlyGlnTrpserApmleuProthnCyAargIle 2546
Db      6913 TCCGAGCCCAATGTCGAGACGAGCCAGTGGAGAGCCCTTGCTCATCTGCGAGTG 6972
Qy      2547 IleancCythraapProglyHniglulnAsnserValArgGlnValHnAlaserGly 2566
Db      6973 GTGAACGTTCCGATCTCGATTTGTGGAATAATGCAAGTCCGCCACGGGCAACAACCTT 7032
Qy      2567 ProHnargPheSerPheGlyThnThValaserTyArgCyAasnHnaglYpHeTyLeu 2586
Db      7033 CCAGAGATTTCGATGAGCAAGTGTATGATCATGCAAGAGAGGGGTTCTACTTA 7092
Qy      2587 LeuglyThnProValleuSerCyaglIngIlyAspGlyThnTrpAargProArgProGln 2606
Db      7093 CTGGGCTCTTCTGCTGACCTGATGCAAGAGTGGCTTGTGGAGCCGCTCTTACCCAG 7152
Qy      2607 CysleuLeuValSerCyaglYHnIleProGlySerProHnIleSerGlnImetSerGlyAsp 2626
Db      7153 TGCTGCGCTATATCATGTGGGATCTGCGGGTCCCGCTAATGCTGTCTGACGTGAGAA 7212
Qy      2627 SerTyThnValGlyAlaValAlaValArgTySerCysIleGlyValArgThnLeuValGly 2646
Db      7213 TTGTTCATTTGAGGCCACAGTTCAGTACTCTGCAAGGGGGCCAGATTCTCACAGGC 7272
Qy      2647 AsnSerThnArgmetCysGlyLeuAspGlyHnIleTrpHnArgIleSerLeuProHnIleCysSer 2666
Db      7273 AATAGCAAGAGTCTGCAAGAGAGAGTCACTGAGAGTATCCCTCCCATTTGTCA 7332
Qy      2667 GlyThnSerValGlyValCyaglYAspProGlyIleProAlaHnIleGlyIleArgLeuGly 2686
Db      7333 GGAATAGTCTCGATTTTGTGTGATCCAGAGAGCCACAGACATGGGTCTCGCTTGGG 7392
Qy      2687 AspSerPheAspProGlyThnValmetArgPheSerCysGluAlaGlyHnIleValLeuArg 2706
Db      7393 GATGAGTTTAAGACAAAGAGCTTTTGCATTTCTCTGATGAGAGGGCCACAGCTCGG 7452
Qy      2707 GlySerSerGluArgThnArgGlnAlaAsnGlySerTrpSerGlySerGlnProGluCys 2726
Db      7453 GGTTCGACAGCGCACATGCTGTGAATGGTCTGCTGAGAGAGTCCAGCCCTGTGT 7512
Qy      2727 GlyValIleSerCysGlyAsnProGlyThnProserAmlaArgValAlaPheSerAsp 2746
Db      7513 GAGGCCGTGCTCTGTGAAACCTTGACACCCCTTACCAATGGAGATGATCTTCAGACGGAT 7572
Qy      2747 GlyLeuValPheSerSerSerIleValTyrgIuCyAargGluGlyTyTyAlaThnGly 2766
Db      7573 GGAATCTCTTCTTCACACTCTGATCATGCTGTGGAGAGGCTCAAGAACTCGGG 7632
Qy      2767 LeuLeuSerArgHnIleCysSerValAsnGlyThnTrpHnArgIleSerAspProGluCysLeu 2786
Db      7633 CTGATGACGGGACCTGACACGAGACGAGACAGAGAGAGAGCCCTGACTGTACA 7692
Qy      2787 ValIleAsnCyaglYAspProGlyIleProAlaAsnGlyLeuArgGluGlyAsnAspPhe 2806
Db      7693 ATCATGACGTGTGTATCTCGGACACCTGCCCAATGGATCCAGTTTGGAGACGATTC 7752
Qy      2807 ArgTyThnValThnValThnTyrgIuCyValAlaProGlyTyrmetmetGluSerHnArg 2826
Db      7753 ACTTTCAACAAGACCGTGAAGTATCAGTGAACCTGCTACTAGAGAGAGAGAGAGAG 7812
Qy      2827 ValSerValleuSerCysThnValAspArgThnTrpAsnGlyThnTyrsProValCysLys 2846
Db      7813 TCACCAACCAATCCCTGACCAAGAGAGTGAATGAGATCAGACCCGAGCCCTTGGAAA 7872
Qy      2847 AlaLeuMetCysLysProProProLeuIleProAsnGlyLysValValGlySerAspPhe 2866

```

```

Db      7873 GCTGTTCTATGACGCCACCTCCCTCATGTGCAAGAGAAAGTGGAGGGGTACAGTTC 7932
Qy      2867 MetTrpGlySerSerValThnTyrgAlaCysLeuGluGlyTyrgIuLeuSerLeuProAla 2886
Db      7933 CGATGGGGTCCAGACATMACTACAGTTGTGTGATGTGATGATGATGATGATGATGATGAT 7992
Qy      2887 ValPheThnCyaglIngIuAsnGlySerTrpThnGlyGluLeuProGlnCysPheProVal 2906
Db      7993 ATCTGTCTGTGAAGGCGGTGAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8052
Qy      2907 PheCysGlyAspProGlyValProserArgGlyArgArgGluAspArgGlyPheSerTy 2926
Db      8053 TTCTGTGCGATTCAGACATCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8112
Qy      2927 ArgSerSerValSerPheSerCysHnAspProProLeuValleuValGlySerProArg 2946
Db      8113 AAGTCGAGGCTTTCATCCAGTCAAAACCCCATTTGTGTGATGAGGGTTCCTGAGAGAGA 8172
Qy      2947 PheCysGlnSerAspGlyThnTrpSerGlyThnGlnProserCysIleAspProThnLeu 2966
Db      8173 ACCTGCGAGCCGAGTGGATGTGAGTGCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8232
Qy      2967 ThnThnCysAlaAspProGlyValProGlnPheGlyIleGlnAsnSerGlnGlyTy 2986
Db      8233 ACCGCTTGCCCAAGACCCCGGACATCCCACTTTGGATACAGATAGCTGGAAGAGATAC 8292
Qy      2987 GlnValGlySerThnValleuPheArgCysGlnLysGlyTyrgLeuLeuGlnGlySerTh 3006
Db      8293 GAGGTGGAGAGCACTGTGTTCTCAGATGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8352
Qy      3007 ThnArgThnCysLeuProAsnLeuThnTrpSerGlyThnProProAspCysValProHn 3026
Db      8353 ACCGGAGCTGTCTTCCCAACCTCACTGAGTGAATCCAGACAGAGTCAATCCCAT 8412
Qy      3027 HnCyAsnGlnProGluThnProThnHnAlaAsnValGlyAlaLeuAspLeuProSer 3046
Db      8413 GCGTGGCGGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8472
Qy      3047 MetGlyTyThnLeuIle 3052
Db      8473 TTGGGCTACACTTATGTC 8490

RESULT 15
ID AAL49944 standard, cDNA, 6004 BP.
XX AAL49944;
XX AC
XX 10-DEC-2002 (first entry)
XX DT
XX DE Human molecule for disease detection and treatment coding sequence #18.
XX XX
XX KW Human; molecule for disease detection and treatment; MBDT; gene therapy;
XX KW cytoskeletal; antiarteriosclerotic; hepatotropic; anti-HIV; anti-allergic;
XX KW anti-inflammatory; antiaspartic; cerebroprotective; nootropic;
XX KW neuroprotective; antiparkinsonian; cardiac; antiangiogenic; gene; ss.
XX OS Homo sapiens.
XX XX
XX PN MO200270709-A2.
XX PD 12-SEP-2002.
XX XX
XX PF 08-FEB-2002; 2002MO-US0003709.
XX XX
XX PR 09-FEB-2001; 2001US-0268117P.
XX PR 15-FEB-2001; 2001US-0269618P.
XX PR 23-FEB-2001; 2001US-0271118P.
XX PR 07-MAR-2001; 2001US-0274486P.
XX PR 09-MAR-2001; 2001US-0274436P.
XX PR 28-NOV-2001; 2001US-0334229P.
XX PR 01-FEB-2002; 2002US-0353284P.
XX XX

```

(INCY-) INCYTE GENOMICS INC.

PA La1 PG, Baughn MR, Yao MG, Malja NK, Elliot VS, Xu Y;
 XX Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM,
 PI Hatalla AA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT,
 PI Rannkumar J, Griffin JA, Swarnaker A, Azimzal Y, Sapperstein SK,
 PI Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;
 XX WPI: 2002-713453/77.
 DR P-PSDB: AMO19415.
 XX
 PT New human molecules for disease detection and treatment (MDT), useful
 PT for diagnosing, treating and preventing diseases or conditions associated
 PT with the aberrant MDT expression, e.g. cancer, AIDS, asthma, diabetes,
 PT hepatitis.
 XX
 PS Claim 11; Page 174-176; 177pp; English.
 XX
 CC The present invention relates to human proteins and coding sequences of
 CC molecules for disease detection and treatment MDT. The sequences can be
 CC used in the treatment of diseases associated with the decreased
 CC expression or overexpression of MDT, such as cell proliferative (cancer,
 CC atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,
 CC allergies, Addison's disease, asthma), developmental (dwarfism, renal
 CC tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
 CC epilepsy) and cardiovascular (congestive heart failure, myocardial
 CC infarction, angina pectoris) disorders. The present sequence is a coding
 CC sequence of the invention
 SQ Sequence 6004 BP; 1272 A; 1795 C; 1615 G; 1322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0	9077.50	6004	1720	5	8	265	8
Percent Similarity:	86.47%						
Best Local Similarity:	86.22%						
Query Match:	53.44%						

DB: 6

US-10-016-248-2 (1-3104) x AAL49944 (1-6004)

QY 859 AsnSerValThrGlyThrGlnGlyThrLeuLeuSerProAsnProValAsnTyrAsn 878
 DB 75 AATTCAGTCAAGACTCA--GGACTTG-CTGTCCCAAGCT-CCCGAATCTCAAT 129
 QY 879 AsnAsnHisGluCysIleTyrSerIleGlnThrGln-Pro-GlyLysGlyIleGlnLeu 898
 DB 130 GACAAATCCTGAATGAT-TACTCCATGCAAGCCAGTCCAGGTAAGCAATTCAGCTCA 188
 QY 898 ValAlaArgAlaPheGluLeuSerGluGlyAspValLeuLysValTyrAspGlyAsnAsn 918
 DB 189 AACCCAGGGGCAATTCGAACTCCGAAGAGAGATGCTCAAGGTTTGTATGGCAACACA 248
 QY 918 AsnSerAlaArgLeuLeuGlyValPheSerHisSerGluMetGlyValThrLeuAsn 938
 DB 249 ACTCCCGCCGTTGCTGGAGAGTTTAAAGCATTTCAAGATGAGGGGAGACTTTGAACA 308
 QY 938 erThiSerSerSerLeuTyrPheAspPheIleThrAspAlaGluAsnThrSerLysGlyP 958
 DB 309 GCACATCCAGCAGCTGTGGCTTGAATTCATCTGATGCTGAAAACAACAGAGAGGCT 368
 QY 958 heGluLeuHisPheSerSerPheGluLeuLysGlyCysGluAspProGlyThrProLysP 978
 DB 369 TTGAATCGCACTTTTCCAGCTTGAACATCAAAAGTAGAGACCCAGAACCCCAAGT 428
 QY 978 heGlyTyrLysValHisAspGluGlyHisPheAlaGlySerSerValSerPheAspCys 998
 DB 429 TTGGCTACAAAGTTCTATGATGAAGTCATTTTCCAGAGAGCTTCGTTCTTCAAGTGG 488
 QY 998 spProGlyTyrSerLeuArgLysSerGluGluLeuLeuCysLeuSerGlyGluArgTyr 1018
 DB 489 ACCCTGGATACAGCCTGGGGGATGAGAGAGCTGCTGTCTGATGGAGAGCCCGGGA 548

QY 1018 hTrpAspArgProLeuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluV 1038
 DB 549 CCGGAGACCGGCTCTGCCACCCTGTGTGCGGAGGTGAGAGACAGAGAGAGAGAG 608
 QY 1038 alSerGlyGlnValLeuSerProGlyTyrProAlaProTyrGlnHisAsnLeuAsnCysI 1058
 DB 609 TGTCGGGGGAGAGTGTGTACACCCGGGTATCCAGCTCCCTATGAAACAAATTCACACTGCA 668
 QY 1058 leTrpThrIleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspT 1078
 DB 669 TCTGGACCATTCAGAGAGAGAGCGCGGTGACCATTTGGGCTACACTTCTGTGTGGACA 728
 QY 1078 heGluGluValHisAspValLeuArgGlyLeTrpAspGlyProValGluSerGlyValLeu 1098
 DB 729 CAGAGAGGTTCACAGAGCTGTGCGCATTTGGAGTGGAGTGGCTTGGAGAGCGGGGTTCTGC 788
 QY 1098 euLysGluLeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerVal 1118
 DB 789 TGAAGAGCTGAGTGGCCCGGCTGTGCCAAGAGACTGTATGACACTTCAACTCGGTGC 848
 QY 1118 alLeuGlnPheSerThrAspPhePheThrSerLysGlyGlyPheAlaIleGlnPheSerV 1138
 DB 849 TCTTCGAGTTCCAGCACTGACTTCTTCCACAGCAGACAGGCGCTTTCATTTTTCAG 908
 QY 1138 alSerThrAlaThrSerCysAsnAspProGlyTyrProGlnAsnGlySerArgSerGly 1158
 DB 909 TGTTCACAGCAACGTCCTGTCAATGACCTTGGGATCCCGAGATGAGATCGGAGTGGTG 968
 QY 1158 spSerTyrPheGluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGln 1178
 DB 969 ACAGTTGGGAGAGCCGGGCACTCCACAGTGTTCAGTGTGACCTGTGCTACGGCTGACAG 1028
 QY 1178 lYSerAlaGlnIleSerCysValLysIleGluAsnArgPhePheThrGlnProSerProP 1198
 DB 1029 GAAAGTCAGAGATCAGCTGTGTGAAGATGAGAAAGGTTCTTCTGGAGAGCCCAAGCCGCG 1088
 QY 1198 roThrCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerP 1218
 DB 1089 CAACATGATGCTGCTCCCTGGGGGAGACCTGACAGACCAATCTGAGATCATCTCTCAC 1148
 QY 1218 roAsnTyrProGluProTyrProProGlyLysGluCysAspTyrLysValThrValSerP 1238
 DB 1149 CAATATCCAGAACCCATACCCGCGCAGGAGAGTGTGATGAGAAATGACCGCTCAC 1208
 QY 1238 roAspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPhe 1258
 DB 1209 CAGACTACGATGCTCCCTGGGTATTTAAATCTTTAACTGAGCTGGCTATGACTTCC 1268
 QY 1258 euHisIleTyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerG 1278
 DB 1269 TCCATATCTACAGAGAGAGGAGCTCTCAAGCCCTCTCATAGGAAGCTTCTATGGCTCC 1328
 QY 1278 lLeuProGlyValArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAsp 1298
 DB 1329 AGCTCCAGGGCGGATGAAAGAGAGCAACACACTTCTTCCCTCCGAGCGATG 1388
 QY 1298 lAserValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGlyLeuSerCysP 1318
 DB 1389 CATCTGTAGAGAAAGCTGTGCTCTGTCACTATGAGAAACCCCGGAGAGCAATGTT 1448
 QY 1318 heAspProGlySerIleLysAsnGlyThrArgValGlySerAspLeuLysLeuGlySers 1338
 DB 1449 TTGATCTCGGTTCCATCAAGAGCGGACACAGGGGTGGGTCCGACTGAAGCTGGGCTCCT 1508
 QY 1338 erValThrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysI 1358
 DB 1509 CCGTCACCTACTACTGCGACAGGGGGCTTACAGAAAGTTGAGGGGACCTTCAGACCTGAGCGCA 1568
 QY 1358 lLeuGlyProAspGlyLysProValTyrAsnAsnProAspProValCysThrAlaProC 1378
 DB 1569 TCCCTGGGGCGCTGAGAGAGAGCCCTGTGTGAAACAATCCCGGCAAGTGTGACAGCCCTCT 1628
 QY 1378 ysGlyGlyGlnTyrValGlySerAspGlyValValLeuSerProAsnTyrProGlnAsnT 1398

```

Db      1629 GTGGGGGACGATAGTGGGTTCCGACGAGGTGCTTGTCCCAACCACTACCCCAAGACT 1688
Qy      1398 YRTHSRGLYGLNILEYSELEUTYRPHVAIPROIALALEULASERTHSERA 1418
Db      1689 ACACGAGTGGACAGATCTGCTGTGATATTTGTAAGTGGCCCAAGACTATGTGGGTTTG 1748
Qy      1418 LYGINPHEALAPHEHISTHRALEUANAAPVALIGLVALHISAPGLHIS 1438
Db      1749 GCCAGTTCGCTTCTTTCACAGGCCCTCAACGAGTGTGAGGTTACAGACGACCA 1808
Qy      1438 ERGINHISERARILEUUSESERLEUSERGLYSERHIS 1452
Db      1809 GCCAGCACTGGCGGTCTCTAGCTCCCTCGGGGTCCATACAGGTATCGGGGGCTCG 1868
Qy      1452 ----- 1452
Db      1869 CCAGTGTGGGATGTTGTGGGCGGGGACATCACTCGGCTAAAGGAGAGGCTCTA 1928
Qy      1452 ----- 1452
Db      1929 GAAGACCCCATGGCCCGCAGGTGGAACCTTAGCGCTCGTCCGTCTGTCTGTG 1988
Qy      1452 ----- 1452
Db      1989 CTGTCTACAGAGCAGACGACGCTCTGAGAGCTCCAACTAGCGGGGCTTCAGCAGCT 2048
Qy      1452 ----- 1452
Db      2049 GCCCTCAACGACTGTGTCTACACCGCCCTTGTGTGAGCTTCTCTGTGATG 2108
Qy      1452 ----- 1452
Db      2109 GCAACTACACTAATTGGCTGACAGTCCAGTTGTGTCTCTCCCTGCGCCACTGTGA 2168
Qy      1452 ----- 1452
Db      2169 CTGCACCAAGAGAAGATATACCTTGTCTCTGCTCAAAAGCTGACGTACCTCG 2228
Qy      1453 ----- 1453
Db      2229 TTTCCTGTGCCATGACGAGGAAATCACTGCTTGGCCCACTCAATCAATTTCTCATTA 2288
Qy      1467 YSPHESERALAYEGLYLEUVALPROIALAARGLYPHENHISPHVALTYRGLN----- 1484
Db      2289 AGTTCAGGGCCAAAGGCTCGCACACGACGAGGCTTCACCTTGTCTACCAAGGTATGG 2348
Qy      1485 ----- 1485
Db      2349 AGGACATGAGACGCCGAGCGGTTCTCTGAAACGACGACGACGACGCTCTGTGCGG 2408
Qy      1499 LUPROAARGTYRGLYLYAARGLEUGLYSERAPPHESERVALGILVALILEVALARGPHEG 1519
Db      2409 AACCCCTCTATGGCAGAGGCTGGGAGTGACTTCTCGGTGGGGCCATCGTCCGCTTGG 2468
Qy      1519 LUCYASANSERGLYTYRALALEUINGLYSERPROGLIILEGLIUCYLEUPROVALPROG 1539
Db      2469 AATCAACTCGGCTATAGCCTCGACGGGGGTGCCAGAGATGAGTGTGCTCTGTGCTG 2528
Qy      1539 LVALALEUALAGLITRPAANVALSERALPROTHRCYVALVALPROCYSEGLYGLYASNT 1559
Db      2529 GGGCTTGGCCCATGGAATCTCTCAGGCCCACTGTGTGTCTCGGTGAGGACCAAC 2588
Qy      1559 EUTHRGILUARGYGLYTHRIIELEUSERPROGLYPHEPROGLUPROTYRIEUSERT 1579
Db      2589 TCACAGAGCGCAGGGGACCAATCTCTGTCTCCCTGCTCCAGACCCCTAACAAGCC 2648
Qy      1579 EUANCYVALITRPLYELLIVALLVALPROGLIUGLYALAGLYILEGLIILEGLINVALS 1599
Db      2649 TCAACTGTGTGGAAGTCTGTGTCTCCGAAAGGCGCTGCATCAAGATTCAAATTGTCA 2708
Qy      1599 ERPHVALITRGLIULASANTRPAAPERLEUGLIVALPHEASPGLYVALASAPANTHYR 1619

```

```

Db      2709 GTTTTGCAGACAGACAACTGGGACTCGCTGGAAGATTGATGTGACATAACTG 2768
Qy      1619 ALTHRMETLEUGLYSERPHESERGLYTHTRVALPROIALALEULASERTHSERA 1639
Db      2769 TAACTAGCTGGGGAGTTTCTCAGAGAACACCGGCTGCGCTTCTGAAACGACCTCCA 2828
Qy      1639 SNGINLEUTYRLEUHIAPHEHYSERAPHISERVALSERALALAGLYPHENHISLEUG 1659
Db      2829 ACCAGCTTACTTCTCATTTTCTACTCAGATATACGCTATCTGACCTGGCTTCCACTTGG 2888
Qy      1659 LUTYRILYSTRVALIGLYLEUSERSEYSPROGLIUPROIALAVALPROSERASNGLYVAL 1679
Db      2889 AGTCAAAACGATGGGCTGAGCAGTTGTCGGAACTGCTGTGTCACAGTAACGGGGTGA 2948
Qy      1679 YSTRHGLIULARGTYRILEUVALASMBAPVALSERPHEGLINCYEGLUPROGLYTYRA 1699
Db      2949 AGACTGGCAGCGCTACTGTGTGATGATGTGTCTTTCACGTGTGAGCCGGGATATG 3008
Qy      1699 LALEUGINGLYHISALAHISILESERCYSMETPROGLYTHTRVALARGTRPASNTRYR 1719
Db      3009 CCTTCAGGGCCAGCCCACTCTCTGCAATGCCGGAACAGTGGCGATGGAATCAC 3068
Qy      1719 TOPROPOLEUCYALIEALAGINCYSGLYGLYTHTRVALIGLULMETGLUGLYVALILEL 1739
Db      3069 CTCTCCACTCTGTATTGACAGATGTGGGGAAACAGTGGAGAGATGAGGGGGTGTATCC 3128
Qy      1739 EUSERPROGLYPHEPROGLYASNTYRPROSERASMETAPCYSETRPLYSILEALAL 1759
Db      3129 TGAAGCCCCGCTTCCAGGCAACTACCCAGTAAACAGACTGCTCTCGAATAATACAC 3188
Qy      1759 EUPROVALIGLYPHEGLYALAHISILEGLINPHEUANAHPHESERTHGLUPROASNHISA 1779
Db      3189 TGCCCGTGGCTTTGAGCTCAATCCAGTCTCTGAACTTCTCCACCGAGCCCAACACG 3248
Qy      1779 SPYRIIEGLIULIEXRASNGLYPROTYRGLUTHRSERATGMETGLYARGPHESERG 1799
Db      3249 ACTACATAGAAATCCGGAATGGCCCTATGAGACAGCCGCAATGATGGAAGATTGAGT 3308
Qy      1799 LYSERGLULEUPROSERSELEULEUSERTHRSERHISGLIUTHTRVALITYRPHENHIS 1819
Db      3309 GAAGCGAGCTTCCAAAGCTCTCTCTTCACAGTCCACGAGACACGCGTATTTCCACA 3368
Qy      1819 ERASPHISERGLINASARGLYPROGLYPHELYSEUGIUTYRGLN----- 1833
Db      3369 GCGACACTCCAGAAATCGCGCAGATTCAAGCTGAGATCAAGATTGACTTACTCCC 3428
Qy      1833 ----- 1833
Db      3429 ACCGATTCTCTCTCTGAGAGTTTGTATCTCTCGGAGTTGGAAGAACAACACTCA 3488
Qy      1834 ----- 1834
Db      3489 CTCCTCCGTCGCGGCTTCTATGTCTGAGATCTTGATCTCTGTTGGAAGCCATGAA 3548
Qy      1837 EUGINGIUCYSPROASERPROGLIUPROPHALIASNGLYILEVALARGLYVALAGLYTYRA 1857
Db      3549 TTCAGAGTCCCAAGACCCAGGCTTCTTCAGATGATGAGGGAGAGCTGGGTGAC 3608
Qy      1857 SVALIGLYINSERVALTHPHEGLIUCYSEUPROGLYTYRGLNLEUTHGLYHISPROV 1877
Db      3609 ACGTGGGACATACGTACTCTTGAAGTCTCCCGGGGTATCAATTGACGCGCACCTG 3668
Qy      1877 ALLEUTHRCYSGINHISGLYTHASARGASNTRPAAPHISPROLEUPROLYSCYSGIUV 1897
Db      3669 TCTCTACGTGTCAACATGACCAACCGGAATCGGACCAACCCCTGCCCCAAGGTGTGAG 3728
Qy      1897 ALPROCYSGIUGLYASNTLETHSERSEASNGLYTHTRVALTYRSEPROGLYPHAPROS 1917
Db      3729 TCCCTTGTGGGGAATACATTTCTTCAACGCACTGTGTACTCCCGGGGTCTCTTA 3788
Qy      1917 ERPROTYRSESERSEGLINASPCEVALITRPLEULIETHTRVAL--PROIIEGLYHIS-G 1936
Db      3789 GCCGTAATCTCACTCCAGACGCTGTCTGTGATCACCGTGGCCAAATTGGCCATGG 3848

```


QY 1936 |yValArgLeuAsnLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleT 1956
 Db 3849 GCGTCCGCTCAACCTGAGCTGCTGAGACAGAGCCCTCTGGAGATTTCATCCACTCT 3908
 QY 1956 rpaapGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLysL 1976
 Db 3909 GGGATGGGCCACACAAACAGCACCGGCTCGGCGCTTTCACCCGAGCATGGCCAGA 3968
 QY 1976 yethrValGlnSerSerSerSerSerGlnValLeuLeuLysPheHisArgAspAlaLathrg 1996
 Db 3969 AAACAGTCAGAGTTCTTCCACAGAGTCTCTGCTCAAGTTCCACGGATGACGCCAG 4028
 QY 1996 |yGlyIlePheAlaIleAlaPheSerAlaTyProLeuThrLysCySPProProThrT 2016
 Db 4029 GGGGAGCTTCGCGCATAGCTTCTCCGCTTATTCACCTACCAAAATGCGCTCTCCACCA 4088
 QY 2016 |leuProAsnAlaGlnValValThrGluAsnGluGluPheAsnIleGlyAspIleValA 2036
 Db 4089 TCTTCCCAACGCGCAAGTCGTCAAGAGATGAAATTCAAATATAGTGACATCGTAC 4148
 QY 2036 rglTyArgCyLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuG 2056
 Db 4149 GCTACAGATGCTCTCCGCTTACTTACTTGGGGAAATGAATTTGACCTGCAAACTTG 4208
 QY 2056 |yThrTyLeuGlnPheGluGlyProProProIleCysGluValHisCysProThrAsnG 2076
 Db 4209 GAACCTACCTGCACTTGAAGGACCAACCCCGATATGTGAAGTGAAGTCTGCAACAAAG 4268
 QY 2076 |leuLeuThrAspSerThrGlyValIleLeuSerGlnSerTyProProGlySerTyProG 2096
 Db 4269 AGCTTCTGACAGACTCCACAGGGGTGATCTGAGCCAGAGCTACCTGGAAAGTATCCCC 4328
 QY 2096 |npheGlnThrCysSerTrpLeuValArgValGluProAspTyAsnIleSerLeuThrV 2116
 Db 4329 AGTTCCAGACCTGCTGTGGCTGTGAGTGAAGGCCCACTTAACATCTCCCTCAGAG 4388
 QY 2116 |alGlnTyPheLeuSerSerGluLysGlnTyAspGluPheGluIlePheAspGlyProSerG 2136
 Db 4389 TGAAGTACTTCCACAGGAGAACCAATATGATGATTGATGATTTGATGGCCATCAG 4448
 QY 2136 |yGlnSerProLeuLeuValAlaLeuSerGlyAsnTySerAlaProLeuIleValThs 2156
 Db 4449 GACAGAGTCTCTCTGTAAGAACCTCAGTGGGAATTACTAGCTCCCTGATTTGTACCA 4508
 QY 2156 erSerSerAsnSerValTyLeuArgTyPheSerSerAspHisAlaTyAsnArgLysGlyP 2176
 Db 4509 GCTCAAGCAACTCTGTGTAAGTCTGTTGGTCACTGATCAAGCTTCAATCCGAAGGCT 4568
 QY 2176 helysIleArgTySerAlaProTyCysSerLeuProArgAlaProLeuHisGlyPheT 2196
 Db 4569 TCAGATTCGCTATTC----- 4585
 QY 2196 |leuGlnGlnThrSerThrGlnProGlyGlySerIleHisPheGlyCysAsnAlaGlyT 2216
 Db 4586 -----GGCCAGACCAACCCAGCCGGGGCTCCATTCACCTTGGCTCAACGCCGCT 4640
 QY 2216 yArgLeuValGlyHisSerMetAlaIleCysThrArgHisProGlnGlyTyHisLLeuT 2236
 Db 4641 ACCGCTGTGGGACACAGCATGGCATGTATCCGGGCAACCCGAGGCTACCACTGT 4700
 QY 2236 rSerSerGlnAlaIleProLeuCysGlnAlaLeuSerCysGlyLeuProGlyAlaProLysA 2256
 Db 4701 GAGAGCAAGCATCTCTCTGTCAAGCTTCTCTGTGGGCTTCTCGTGGGCCCCCAAGA 4760
 QY 2256 snGlyMetValPheGlyLysGlyTyThrValGlyTyThrLysAlaValTySerCysSerG 2276
 Db 4761 ATGGAATGGTGTGGCAGAGTACAGAGTGAAGCAAGCCCATGTACACTGCTCAGTG 4820
 QY 2276 |uGlyTyHisLLeuGlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuT 2296
 Db 4821 AAGGCTACCACTCCAGGAGGCGGTGAAGGCACTGCAAGATGTCTGAGCAACAGGCTAT 4880

QY 2296 rSerAsnArgAsnValProProGlnCysVal-----P 2307
 Db 4881 GAGAGCAACCGCAATGTCCACACAGGTGTGCTCGAGTCTCGGGCAATGAGAGCGGT 4940
 QY 2307 roValThrCysProAspValSerSerIleSerValGluHisGlyArgTrpArgLeuIleP 2327
 Db 4941 CTGTGACTTGTCTGTATGTCAATAGCATCAGGTGGAGCATGGCCGATGGAGCTTATCT 5000
 QY 2327 heGlnThrGlnTyGlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTyTyTyT 2347
 Db 5001 TTGAGACACAGTATCAGTTCCAGGCCAGCTGATCTCATCTGTGACCTCTGCTACT 5060
 QY 2347 yThrGlyGlnArgValIleArgCysGlnAlaAsnGlyLysTrpSerLeuGlyAspSerT 2367
 Db 5061 ATACTGGCCAAAGGTATCTCCGTGACGCCCAATGGCAATGAGACCTCGGGAGCTTGA 5120
 QY 2367 hProThrCysArgIleIleSerCysGlyLysProIleProProAsnGlyHisArgI 2387
 Db 5121 CCGCCACCTGCGGAATCATCTCTGTGGAGAGCTCCCGATTCCCCCAATGGCCACCGCA 5180
 QY 2387 |eGlyThrLeuSerValTyArgValaThrAlaIlePheSerCysAsnSerGlyTyThrL 2407
 Db 5181 TCGAAGCACTGTCTGTCTACGGGGCAACAGCCATCTTCTCGCAATTCGGATCACAC 5240
 QY 2407 euValGlySerArgValArgGluCysMetAlaAsnGlyLysThrProSerGlySerGluValA 2427
 Db 5241 TGTGGGGCTCCAGGGTGGGTGATGTCAATGGCCAAATGGGCTCTGAGTGGCTTGAAGTCC 5300
 QY 2427 rGlyLeu----- 2429
 Db 5301 GCTGCTTGGCCACTCAGACCAAGCTCACTCCATTTCTATAGCTCTCTTCATGTAC 5360
 QY 2430 -----AlaGlyHisCysGlyTyThrProGluProIleValAsn 2442
 Db 5361 TCTCTTCCCATCTCCCTCAACAAAGCTGACACTGTGGAGCTCTCGAGCCATTTGCAACG 5420
 QY 2442 |yHisIleAsnGlyGluAsnTySerTyArgGlySerValValTyGlnCysAsnAlaG 2462
 Db 5421 GACACATCAATGGGGAGAACTACAGCTACCGGGGAGTGTGTATCCAAATGCATGTCTG 5480
 QY 2462 |yPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTrpSerGlyL 2482
 Db 5481 GCTTCCGCTGATGGCATGTCTGTGCGCATCTGCACAGAGATCATCATCTGTCGGCA 5540
 QY 2482 yThrProPheCysVal----- 2487
 Db 5541 AGACCCCTTCTGTGTGATGTAAAGACAGCTGCTGCTGCTGCTGTGTGTG 5600
 QY 2488 -----ProIleThrCysGlyHisProGlyAsn 2497
 Db 5601 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5660
 QY 2497 roValAsnGlyLeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValValLysPheValC 2517
 Db 5661 CTGTCAAGGCTCACTCAGGGTAAACAGTTTAACTTCAAGAGATGTGTCAAGTTTGT 5720
 QY 2517 yAsnProGlyTyMetAlaGluGlyAlaIleArgSerGlnCysLeuAlaSerGlyGlnT 2537
 Db 5721 GCAACCTGGGTATATAGCTGAGGGGCTGTAGTCCCAATGCTGCGCCAGCGGCAAT 5780
 QY 2537 rSerAspMetLeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGluA 2557
 Db 5781 GAGATGACATGCTGCCACCTGCGAATCATCACTATACAGATCTCGACACCAAGAAA 5840
 QY 2557 snSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyTyThrValS 2577
 Db 5841 ATAGTGTCTCTCAGGTCCAGCCAGCGGCCGACAGGTTCAGTTTGGGACCACTGTGT 5900
 QY 2577 erTyArgCysAsnHisGlyPheTyLeuLeuGlyThrProValLeuSerCysGlnGlyA 2597
 Db 5901 CTTACCGGTGCAACCAAGGCTTACTCTGTGGGCAACCCAGTGTCTCAGTGTCCAGGGAG 5960
 QY 2597 spGlyThrTrpAspArgProArgProGlnCysLeu 2608

Db 5961 ATGGCACATGGGACCGTCGCCGCCCGCAGTGTCTC 5995

Search completed: October 18, 2004, 16:07:48
Job time : 4182 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 09:58:21 ; Search time 150 Seconds

(Without alignments)
11906.419 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPPALLPCLSLSDCC.....RSGVDPSTLPGRSPK 3104

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0 %
Maximum Match 100 %

Listing first 45 summaries

Database :

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15364.5	90.5	3487	1 CSM2 HUMAN	Q7408 homo sapien
2	11372.5	67.0	3565	1 CSM1 HUMAN	Q96P27 homo sapien
3	11305	66.6	3564	1 CSM1 MOUSE	Q92313 mus musculu
4	10591.5	62.4	3670	1 CSM3 HUMAN	Q7407 homo sapien
5	8997.5	53.0	2796	1 CSM3 MOUSE	Q80C79 mus musculu
6	6180.5	36.4	1466	2 Q72241	Q72241 brachydantio
7	1864.5	11.0	3567	2 Q96S77	Q96S77 mus musculu
8	1550	9.1	2489	2 Q16744	Q16744 homo sapien
9	1473	8.7	3494	2 Q7LC53	Q7LC53 homo sapien
10	1473	8.7	3623	2 Q60494	Q60494 homo sapien
11	1457.5	8.6	3620	2 Q9TUS3	Q9TUS3 canis famli
12	1416	8.3	1441	2 Q723G3	Q723G3 homo sapien
13	1415	8.3	1497	2 Q8NBT9	Q8NBT9 homo sapien
14	1404.5	8.3	3623	2 Q70244	Q70244 rattus norv
15	1390	8.2	2014	2 Q29530	Q29530 pan troglod
16	1372	8.1	1323	2 Q72387	Q72387 homo sapien
17	1371	8.1	2039	1 CR1 HUMAN	P17927 homo sapien
18	1298	7.6	1911	2 Q29528	Q29528 papio hamad
19	1190	7.0	3646	2 Q70737	Q70737 anophelae g
20	968.5	5.7	3687	2 Q9W333	Q9W333 drosophila
21	906.5	5.3	1025	1 CR2 MOUSE	P19070 mus musculu
22	906.5	5.3	1032	2 Q9DC83	Q9DC83 mus musculu
23	906	5.2	1174	2 Q9VYR4	Q9VYR4 drosophila
24	889.5	5.2	1033	1 CR2 HUMAN	P20023 homo sapien
25	878.5	5.2	1135	2 Q70137	Q70137 anophelae g
26	864.5	5.1	974	2 P91658	P91658 drosophila
27	832	4.9	1045	2 Q46545	Q46545 ovae aries
28	797	4.7	869	2 Q922H0	Q922H0 mus musculu
29	793	4.7	869	2 Q8NB57	Q8NB57 homo sapien
30	793	4.7	972	2 Q8ND50	Q8ND50 homo sapien
31	788	4.6	996	2 Q8TD25	Q8TD25 homo sapien

Result 1	ID	CSM2 HUMAN	STANDARD	PRT	3487 AA	057460 brachydantio
AC	Q72408	Q8N963	Q96C03	Q9H4V7	Q9H4V8	Q9H4V9
AC	Q9H4M2	Q9H4W3	Q9H4W4	Q9HCY5	Q9HCY6	Q9HCY7
DT	29-MAR-2004	(Rel. 43, Created)				
DT	29-MAR-2004	(Rel. 43, Last sequence update)				
DT	05-JUL-2004	(Rel. 44, Last annotation update)				
DE	CUB and sushi multiple domains protein 2.					
GN	Name=CSMD2; Synonyms=KIAA1884;					
OS	Homo sapiens (Human).					
OC	Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.					
RX	MEHLIN-22788796; PubMed=12906867; DOI=10.1016/S0888-7543(03)00149-6;					
RA	Lau W.L., Scholnick S.B.;					
RT	"Identification of two new members of the CSMD gene family.";					
RL	Genomics 82:412-415(2003).					
RN	[2]					
RP	SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 2969-3487 (ISOFORM 1).					
RC	TISSUE=Brain, and Testis; PubMed=14702039; DOI=10.1038/ng1285;					
RX	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosogi T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiyaoka S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Horiuchi T., Kusanagi J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Mutsaers K., Yuki H., Oshima A., Sasaki N., Aoshima S., Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujisawa T., Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Okabe A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsunaga K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Tashiro T., Oyama M., Hata H., Watanabe M., Komatsu T., Nakagawa K., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;					
RT	"Complete sequencing and characterization of 21,243 full-length human cDNAs.";					

ALIGNMENTS

32	785.5	4.6	1022	1	TLD BRARE	057460 brachydantio
33	772.5	4.5	1008	2	Q9DER7	Q9DER7 gallus gall
34	771.5	4.5	1012	2	Q7SK2	Q7SK2 mus musculu
35	771	4.5	1012	2	Q9WVW6	Q9WVW6 mus musculu
36	769.5	4.5	1013	2	Q9NOS4	Q9NOS4 homo sapien
37	768.5	4.5	1013	2	Q43897	Q43897 homo sapien
38	764	4.5	991	2	Q62223	Q62223 mus musculu
39	763.5	4.5	977	2	Q62269	Q62269 mus musculu
40	763	4.5	775	2	Q6P550	Q6P550 mus musculu
41	763	4.5	775	2	AAH63079	AAH63079 mus muscu
42	763	4.5	991	2	Q6NZM2	Q6NZM2 mus musculu
43	763	4.5	991	2	AAH66062	AAH66062 mus muscu
44	762.5	4.5	1007	2	Q8J128	Q8J128 xenopus lae
45	761	4.5	1024	1	SZ6L_HUMAN	Q9BYH1 homo sapien

RL Nat. Genet. 36:40-45(2004).
 RN (3)
 RP SEQUENCE FROM N.A.
 RA Wallis J., Brown A.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stalcup M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci F., Prange C.,
 RA Bork S.A., McEwan P.J., McKernan K.J., Abramson R.D., Mullenbach S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Hellon E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP (5)
 RP SEQUENCE OF 2542-3487 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=21456161; PubMed=11572484;
 RA Nagase T., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXI.
 RT The complete sequences of 60 new cDNA clones from brain which code for
 RT large proteins.";
 RL DNA Res. 8:179-187(2001).
 RL -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q72408-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q72408-2; Sequence=VSP_009038, VSP_009039, VSP_009040,
 CC VSP_009041, VSP_009042;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q72408-3; Sequence=VSP_009043, VSP_009044, VSP_009045,
 CC VSP_009046;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in
 CC brain. Expressed at intermediate level in brain, including
 CC cerebellum, substantia nigra, hippocampus and fetal brain.
 CC Overexpressed in some head and neck cancer cell lines.
 CC -1- SIMILARITY: Belongs to the CSMD family.
 CC -1- SIMILARITY: Contains 14 CUB domains.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 939 that shortens the protein by 1021
 CC residues in its N-terminus. It is unknown whether the sequence
 CC shown exists or whether Ref.3 is right, shortening the sequence in
 CC its N-terminus.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AY10418; AAC34701.1; -;
 DR EMBL; AK095627; BAC04593.1; ALT_INIT.
 DR EMBL; AK127722; BAC87101.1; -;
 DR EMBL; AC115285; -; NOT_ANNOTATED_CDS.
 DR EMBL; AC115286; -; NOT_ANNOTATED_CDS.
 DR EMBL; AL121980; CAC05319.1; ALT_SEQ.
 DR EMBL; AL121980; CAC05320.1; ALT_SEQ.
 DR EMBL; AL121980; CAC05321.1; ALT_SEQ.
 DR EMBL; AL121980; CAC05322.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10283.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10284.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10285.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10286.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10287.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10288.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10289.1; ALT_SEQ.
 DR EMBL; AL355178; CAC10290.1; ALT_SEQ.
 DR EMBL; AL596224; -; NOT_ANNOTATED_CDS.
 DR EMBL; AL607106; -; NOT_ANNOTATED_CDS.
 DR EMBL; BC031871; AH31871.1; -;
 DR EMBL; AB067471; BAB67777.1; -;
 DR HSSP; P10998; 1VD.
 DR Genew; HGNC:19290; CSMD2.
 DR MIM; 608398; -;
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00431; CUB; 14.
 DR Pfam; PF00084; Sushi; 26.
 DR PROSITE; PS01180; CUB; 14.
 DR PROSITE; PS0923; SUSHI; 26.
 KW Alternative splicing; Direct protein sequencing; Repeat; Sushi;
 KW Transmembrane.
 FT DOMAIN 1 3408 Extracellular (Potential).
 FT TRANSMEM 3409 3429 Potential.
 FT DOMAIN 3430 3487 Cytoplasmic (Potential).
 FT DOMAIN 26 134 CUB 1.
 FT DOMAIN 137 198 Sushi 1.
 FT DOMAIN 202 306 Sushi 2.
 FT DOMAIN 341 402 Sushi 2.
 FT DOMAIN 405 516 Sushi 3.
 FT DOMAIN 519 576 Sushi 3.
 FT DOMAIN 578 686 Sushi 4.
 FT DOMAIN 689 750 Sushi 4.
 FT DOMAIN 752 860 Sushi 5.
 FT DOMAIN 865 922 Sushi 5.
 FT DOMAIN 924 1034 CUB 6.
 FT DOMAIN 1037 1096 Sushi 6.
 FT DOMAIN 1098 1206 CUB 7.
 FT DOMAIN 1209 1269 Sushi 7.
 FT DOMAIN 1271 1380 CUB 8.
 FT DOMAIN 1383 1443 Sushi 8.
 FT DOMAIN 1445 1553 CUB 9.
 FT DOMAIN 1556 1617 Sushi 9.
 FT DOMAIN 1619 1727 CUB 10.
 FT DOMAIN 1733 1794 Sushi 10.
 FT DOMAIN 1796 1904 CUB 11.
 FT DOMAIN 1907 1966 Sushi 11.
 FT DOMAIN 1968 2076 CUB 12.
 FT DOMAIN 2121 2180 Sushi 12.
 FT DOMAIN 2182 2293 CUB 13.
 FT DOMAIN 2292 2353 Sushi 13.
 FT DOMAIN 2355 2466 CUB 14.
 FT DOMAIN 2466 2528 Sushi 14.
 FT DOMAIN 2529 2590 Sushi 15.
 FT DOMAIN 2591 2655 Sushi 15.
 FT DOMAIN 2656 2713 Sushi 16.
 FT DOMAIN 2714 2771 Sushi 17.
 FT DOMAIN 2772 2829 Sushi 18.
 FT DOMAIN 2833 2890 Sushi 19.
 FT DOMAIN 2891 2949 Sushi 20.
 FT DOMAIN 2950 3009 Sushi 21.
 FT DOMAIN 3010 3067 Sushi 22.
 FT DOMAIN 3068 3125 Sushi 23.
 FT DOMAIN 3125 Sushi 24.

FT DOMAIN 3129 3187 Sueni 25.
FT DOMAIN 3188 3247 Sueni 26.
Query Match 90.5%; Score 15364.5; DB 1; Length 3487;
Best Local Similarity 91.9%; Pred. No. 0;
Matches 2847; Conservative 12; Mismatches 39; Indels 199; Gaps 10;

2 AGAPPPL-----LLPCSLIDSCASNORHSVGVSESLYKQILEKSGVLTMPKNSQ 57
271 ASLPAPVYSSKMWLRHFTSD--GNHRQR-GFSAQYQVKQILEKSGVLTMPKNSQ 326
58 KTSVLTQVGVSGQNMCPDPCIPEKGRKLGSDFLRGSSVQFTCEGYDQSGKEITCKV 117
327 KTSVLTQVGVSGQNMCPDPCIPEKGRKLGSDFLRGSSVQFTCEGYDQSGKEITCKV 386
118 SDMPAASDHRPVCARMCDALHGPSGITSPNPIQYDNNAHCWITTLNPSKYVL 177
387 SDMPAASDHRPVCARMCDALHGPSGITSPNPIQYDNNAHCWITTLNPSKYVL 446
178 AFEEFDLERGYDTLVGDGDGDKTVLVMQNAQSPHTPSRTPESMGDIWQXW 237
447 AFEEFDLERGYDTLVGDGDGDKTVLVMQNAQSPHTPSRTPESMGDIWQXW 495
238 TVLEICHDISSDARSQSVRSKPTSNVAVETIEQSGCDPGIPAYGREGSRFH 297
496 LLFG-----TDG-SGS-----SLGFKASVEIEQSGCDPGIPAYGREGSRFH 538
298 HGDTLKECPAPFELVGQKAITCOKNQWAKKQCVSCFPNFTSPSGVLSFNPEDY 357
539 HGDTLKECPAPFELVGQKAITCOKNQWAKKQCVSCFPNFTSPSGVLSFNPEDY 598
358 GNHLHCWLLIARPESTRHAFNDIDVEPODFLVIKQATAEAPVLTGFGNDLPSSIT 417
599 GNHLHCWLLIARPESTRHAFNDIDVEPODFLVIKQATAEAPVLTGFGNDLPSSIT 658
418 SSGVAVARLEFQTDHSTGKGFNITFTFRHNECPDPGVAVNGKAFGSDLOQSSISFLCD 477
659 SSGVAVARLEFQTDHSTGKGFNITFTFRHNECPDPGVAVNGKAFGSDLOQSSISFLCD 718
478 EGPLTQSGSETITCVLKEGSVVMNSAVLRCEAPCGGHLTSPSGITLSPGMPGYKDALSC 537
719 EGPLTQSGSETITCVLKEGSVVMNSAVLRCEAPCGGHLTSPSGITLSPGMPGYKDALSC 778
538 AMVLEAGQGYPIKLTDFDRFTEVNYDTLEVBDGRYSAPLIGVHGTQVPOFLISTSYL 597
779 AMVLEAGQGYPIKLTDFDRFTEVNYDTLEVBDGRYSAPLIGVHGTQVPOFLISTSYL 838
598 YLLFSTDKSHDIFGQLEYETITLOSDBCLDPGIPVNGQRHGNDFYVVALVTFSCDSGYT 657
839 YLLFSTDKSHDIFGQLEYETITLOSDBCLDPGIPVNGQRHGNDFYVVALVTFSCDSGYT 898
478 EGPLTQSGSETITCVLKEGSVVMNSAVLRCEAPCGGHLTSPSGITLSPGMPGYKDALSC 537
719 EGPLTQSGSETITCVLKEGSVVMNSAVLRCEAPCGGHLTSPSGITLSPGMPGYKDALSC 778
538 AMVLEAGQGYPIKLTDFDRFTEVNYDTLEVBDGRYSAPLIGVHGTQVPOFLISTSYL 597
779 AMVLEAGQGYPIKLTDFDRFTEVNYDTLEVBDGRYSAPLIGVHGTQVPOFLISTSYL 838
598 YLLFSTDKSHDIFGQLEYETITLOSDBCLDPGIPVNGQRHGNDFYVVALVTFSCDSGYT 657
839 YLLFSTDKSHDIFGQLEYETITLOSDBCLDPGIPVNGQRHGNDFYVVALVTFSCDSGYT 898
658 LSDPELECEBPNFQMSRALPSCCALCGGFIQSSGTTILSPGPDPFPYNNLCTWIIETSH 717
899 LSDPELECEBPNFQMSRALPSCCALCGGFIQSSGTTILSPGPDPFPYNNLCTWIIETSH 958
718 GKGVFTFTHTHLESSGHYLLITENGSTQRLQTLGSRLLPAPISAGIYGNFTAQVRIS 777
959 GKGVFTFTHTHLESSGHYLLITENGSTQRLQTLGSRLLPAPISAGIYGNFTAQVRIS 1018
778 DFSVSYEGFNITFSEYDLEPCBEPVPAYSIRKGLQFVGDTLTFSCPGYRLSGTATIT 837
1019 DFSVSYEGFNITFSEYDLEPCBEPVPAYSIRKGLQFVGDTLTFSCPGYRLSGTATIT 1078
838 CLGGRARLWSSPLPRCAVACGNSVGTGTLSPNFPVNNYNNHCEIYSIQTOPKGIQL 897
1079 CLGGRARLWSSPLPRCAVACGNSVGTGTLSPNFPVNNYNNHCEIYSIQTOPKGIQL 1138
898 KARAFELSEGQVLLKYDDNNNSARLLGVFSHSEMMGVTLLNSTSSLMWDTTDAENSKG 957
1139 KARAFELSEGQVLLKYDDNNNSARLLGVFSHSEMMGVTLLNSTSSLMWDTTDAENSKG 1198
958 FELHFSFELLKCEDPGRPFGYKVHDEGHFAGSSVSFSCPGYSLRGEELLCLSGRR 1017

1199 FELHFSFELLKCEDPGRPFGYKVHDEGHFAGSSVSFSCPGYSLRGEELLCLSGRR 1258
1018 TWBRPLPTVAECGGTTRGEVSGOVLSPGAPVREHNLNCITWTEARAGCTIGHPLVFD 1077
1259 TWBRPLPTVAECGGTTRGEVSGOVLSPGAPVREHNLNCITWTEARAGCTIGHPLVFD 1318
1078 TEEVHDVLRIMDGVESGVLKELSGPALPKDLISTNSVVLQSTDFPFTSKOFAILOFS 1137
1319 TEEVHDVLRIMDGVESGVLKELSGPALPKDLISTNSVVLQSTDFPFTSKOFAILOFS 1378
1138 VSTATSCNDPGIPONGSRSDGSWEAGDSTVPOCPGALQGSABEISCVKIEENRFWQSP 1197
1379 VSTATSCNDPGIPONGSRSDGSWEAGDSTVPOCPGALQGSABEISCVKIEENRFWQSP 1438
1198 PTCIAPCGDLTGSSGVTILSPNYPEPPEKCECMKTVSPDYIALVFNIFNLEPGYDF 1257
1439 PTCIAPCGDLTGSSGVTILSPNYPEPPEKCECMKTVSPDYIALVFNIFNLEPGYDF 1498
1258 LHIYDGRDLSPLIGSFYGSOLPGRIESSNSLFLARSDASVSNAGFVIDYENPRESC 1317
1499 LHIYDGRDLSPLIGSFYGSOLPGRIESSNSLFLARSDASVSNAGFVIDYENPRESC 1558
1318 FDPGSIKNGTRVGSDDLKLGSSVTYYCHGYEVEGTSTLSCILGPDGKPVNNRPVCTAP 1377
1559 FDPGSIKNGTRVGSDDLKLGSSVTYYCHGYEVEGTSTLSCILGPDGKPVNNRPVCTAP 1618
1378 CGGQYVSSDGVVLSPNTPONTSGOICLYPTVYKQYVVGQAFEFHTALNDVYVHDG 1437
1619 CGGQYVSSDGVVLSPNTPONTSGOICLYPTVYKQYVVGQAFEFHTALNDVYVHDG 1678
1438 SOHSRLSSLSGSHTGESLPLATSNQVLIKFSAGKLAPARGFHVQYAVPRTASTQCSSV 1497
1679 SOHSRLSSLSGSHTGESLPLATSNQVLIKFSAGKLAPARGFHVQYAVPRTASTQCSSV 1738
1498 PEPYRGRLGSDSEVGAIVRECEGSGYALQGSPEIECLPVGALAQNNVSAPTCVPCGG 1557
1739 PEPYRGRLGSDSEVGAIVRECEGSGYALQGSPEIECLPVGALAQNNVSAPTCVPCGG 1798
1558 NLTERRGITLSPGPEBYLANSKCVKIIVPEAGIOIOVVSFTTEQNMDSLEVPDQADN 1617
1799 NLTERRGITLSPGPEBYLANSKCVKIIVPEAGIOIOVVSFTTEQNMDSLEVPDQADN 1858
1518 TVTMLGFSFGTTPALNSTSNOLYHFYSDISVSAAGFHELYTVGSSCPPEAVPNSG 1677
1859 TVTMLGFSFGTTPALNSTSNOLYHFYSDISVSAAGFHELYTVGSSCPPEAVPNSG 1918
1678 VKTGERVLYVNDVVSFOCEPGYALQGHAIISCMPTVARMVPPLCTIAOCGGIVEMEGV 1737
1919 VKTGERVLYVNDVVSFOCEPGYALQGHAIISCMPTVARMVPPLCTIAOCGGIVEMEGV 1978
1738 ILSFGPFGNTPSNMDCSWKIALPVFGAHIOPLNFSTEPNHDYIEIRNGPYETSRMGRF 1797
1979 ILSFGPFGNTPSNMDCSWKIALPVFGAHIOPLNFSTEPNHDYIEIRNGPYETSRMGRF 2038
1798 SGESEPLSLISTSHETTVYFHSDHSONRPGFKLEYO----- 1833
2039 SGESEPLSLISTSHETTVYFHSDHSONRPGFKLEYO----- 2098
1834 -----AYELOCPDPEPFANGIVAGAGYNNQSVTFBCLPGYOLTG 1875
2099 STPPVAASYWDDLOPGEAYELOCPDPEPFANGIVAGAGYNNQSVTFBCLPGYOLTG 2158
1876 PVLTCQHGTRNRMDHPLPKCEVPCGNSITSSNGTVSPGPPSSSQDCVWLITVPIGH 1935
2159 PVLTCQHGTRNRMDHPLPKCEVPCGNSITSSNGTVSPGPPSSSQDCVWLITVPIGH 2218
1936 GVRNLNLSLLOTEPSGDIITIMDGPQOTAPRLGVFTRSMAKKTVOSSNOVLLKTHRDAAT 1995
2219 GVRNLNLSLLOTEPSGDIITIMDGPQOTAPRLGVFTRSMAKKTVOSSNOVLLKTHRDAAT 2278
1996 GGIFALFASAVPLTKCPEPTTLPNAAVVTENEENIDIVRYRCLPGFTLVGNELTKL 2055
2279 GGIFALFASAVPLTKCPEPTTLPNAAVVTENEENIDIVRYRCLPGFTLVGNELTKL 2338

QY 2056 GTVLOEGPPIICEVHCPTNELLTDS TGVLSSOSVPGSVPOFOTCSMLVVRPDPVNIET 2115
 Db 2339 GTVLOEGPPIICEVHCPTNELLTDS TGVLSSOSVPGSVPOFOTCSMLVVRPDPVNIET 2398
 QY 2116 VEYFISEKQYDEFEIFDPSGSGPLLKALSGNYSAPLVTSSSSNVLRMSDHAHYNRKG 2175
 Db 2399 VEYFISEKQYDEFEIFDPSGSGPLLKALSGNYSAPLVTSSSSNVLRMSDHAHYNRKG 2458
 QY 2176 FIKRISAPYCSLPRLPIHGFILIGOSTQSGSIHGCNAGRLVGHSMALCTRRPQGYHL 2235
 Db 2459 FIKRISAPYCSLPRLPIHGFILIGOSTQSGSIHGCNAGRLVGHSMALCTRRPQGYHL 2518
 QY 2236 WSEALPLCOLALSCGLPEAPKNGMVEGKEYTVGTRKAVCSCEGYHLQAGABATACLDLQGL 2295
 Db 2519 WSEALPLCOLALSCGLPEAPKNGMVEGKEYTVGTRKAVCSCEGYHLQAGABATACLDLQGL 2578
 QY 2256 WSNRNVPPQCVVTPDVSISVEHGRWRLIFETQYFOAOLMLICDGYYYTGORVYRC 2355
 Db 2579 WSNRNVPPQCVVTPDVSISVEHGRWRLIFETQYFOAOLMLICDGYYYTGORVYRC 2638
 QY 2356 QANGKMSIGDSTPTCRITISCGELPIPNHGRIIGTLVYGATATSCNSGYTLVSSRVREC 2415
 Db 2639 QANGKMSIGDSTPTCRITISCGELPIPNHGRIIGTLVYGATATSCNSGYTLVSSRVREC 2698
 QY 2416 MANGIMSGSEVRCLAGHCCTPEPIVNGHNGENYSSRGSVYQCNAGFRILGMSVRICQ 2475
 Db 2699 MANGIMSGSEVRCLAGHCCTPEPIVNGHNGENYSSRGSVYQCNAGFRILGMSVRICQ 2758
 QY 2476 DHHMSGKTPFCVPIITCGHRPNVANGLOGNPNANDVVKPCNPGYMAEGANRSLASG 2535
 Db 2759 DHHMSGKTPFCVPIITCGHRPNVANGLOGNPNANDVVKPCNPGYMAEGANRSLASG 2770
 QY 2536 QMSDMLPTCRITINCTPDGHNSVROVHASGPRHFSFGTIVSYRCHNGFYTLGTPVLSCQ 2595
 Db 2771 QMSDMLPTCRITINCTPDGHNSVROVHASGPRHFSFGTIVSYRCHNGFYTLGTPVLSCQ 2770
 QY 2556 GDGTWDRRPOCLLVSCGHPSPPHSGMSGDSYTVGAVVARYSCIGKRTLVGNSTRMGDLD 2655
 Db 2771 GDGTWDRRPOCLLVSCGHPSPPHSGMSGDSYTVGAVVARYSCIGKRTLVGNSTRMGDLD 2817
 QY 2656 GHMTGSLPHCSGTSVYCGDPGIPAHGIRLIGDSDPCTWNRPFCEAGHYLRGSSERTCQA 2715
 Db 2818 GHMTGSLPHCSGTSVYCGDPGIPAHGIRLIGDSDPCTWNRPFCEAGHYLRGSSERTCQA 2877
 QY 2716 NGWSGSGQPECGVISCGNPGTSPNARVVFSDGLVFSSSIYEGREGYATGLSRHCSVA 2775
 Db 2878 NGWSGSGQPECGVISCGNPGTSPNARVVFSDGLVFSSSIYEGREGYATGLSRHCSVA 2937
 QY 2776 GTWGSDEPCLVINGCGPBI PANGLRLGNDPFRNKTYYTQCVCYGMESHRSVLSCTD 2835
 Db 2938 GTWGSDEPCLVINGCGPBI PANGLRLGNDPFRNKTYYTQCVCYGMESHRSVLSCTD 2997
 QY 2836 RIMNGTRPVCKALMKCPPLIPNGKVVSGDFMWSGSSTVYACLEGYOLSPAVTCEGNS 2895
 Db 2998 RIMNGTRPVCKALMKCPPLIPNGKVVSGDFMWSGSSTVYACLEGYOLSPAVTCEGNS 3057
 QY 2896 WTEBELPQCFVYFCGDPVBSRGRREDRGSYSRVSFSPCHPLVLVGSRRRRCQSDGWS 2955
 Db 3058 WTEBELPQCFVYFCGDPVBSRGRREDRGSYSRVSFSPCHPLVLVGSRRRRCQSDGWS 3117
 QY 2956 GTOPSCIDPFLTTCADBGVPOFGIIONNSOGYGVSTVLRCKCKGYTLQOSTTRTCLPNT 3015
 Db 3118 GTOPSCIDPFLTTCADBGVPOFGIIONNSOGYGVSTVLRCKCKGYTLQOSTTRTCLPNT 3177
 QY 3016 WSGTPDPCVPHHCRQEPPTTHANVAGLDPMSMGYTLI 3052
 Db 3178 WSGTPDPCVPHHCRQEPPTTHANVAGLDPMSMGYTLI 3214

RESULT 2
 CSM1_HUMAN
 ID_CSM1_HUMAN STANDARD; PRT; 3565 AA.

AC Q96P27; Q96Q09; Q96RM4;
 DT 29-MAR-2004 (Rel. 43, Created)
 DT 29-MAR-2004 (Rel. 43, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE CUB and sushi multiple domains protein 1 precursor (UNQ5952/PRO19863).
 GN Name=CSM1; Synonyms=K1AA1890;
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=21365705; PubMed=1172063; DOI=10.1006/geno.2001.6587;
 RX Sun P.C., Uppaluri R., Schmidt A.P., Pashia M.E., Quant E.C.,
 RA Sunwoo J.B., Gollin S.M., Schoenick S.B.;
 RT "transcript map of the 8p23 putative tumor suppressor region.";
 RL Genomics 75:17-25(2001).
 RN [2]
 RP SEQUENCE OF 1319-3565 FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Iosida M., Huchita T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togliya S., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Imose N., Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
 RA Nakai R., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 RN [3]
 RP SEQUENCE OF 966-2013 FROM N.A. (ISOFORM 4), AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=21456161; PubMed=11572484;
 RA Nagase T., Kikuno R., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
RN [5]
RP DISEASE.
RX PubMed=12696061; DOI=10.1002/gcc.10191;
RA Tomes C., Jackson A., Maguire K., Wood J., Gollin S., Ishwad C.,
RA Paterson I., Prime S., Parkinson K., Bell S., Woods G., Matham A.,
RA Oliver R., Woodward R., Sloan P., Dixon M., Read A., Thakker N.,
RT "The presence of multiple regions of homozygous deletion at the CSDM1
RT locus in oral squamous cell carcinoma question the role of CSDM1 in
RT head and neck carcinogenesis.";
RL Genes Chromosomes Cancer 37:132-140(2003).
RN [6]
RP DISEASE.
RX PubMed=14506705; DOI=10.1002/gcc.10279;
RA Scholnick S.B., Richter T.M.;
RT "The role of CSDM1 in head and neck carcinogenesis.";
RL Genes Chromosomes Cancer 38:281-283(2003).
CC -1- FUNCTION: Potential suppressor of squamous cell carcinomas.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q96P27-1; Sequence=Displayed;
CC Name=2; Synonym=Short;
CC IsoId=Q96P27-2; Sequence=VSP_009034, VSP_009035;
CC Name=3;
CC IsoId=Q96P27-3; Sequence=VSP_009030, VSP_009031;
CC Note=No experimental confirmation available;
CC Name=4;
CC IsoId=Q96P27-4; Sequence=VSP_009032, VSP_009033;
CC -1- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in
CC brain. Expressed at intermediate level in brain, including
CC cerebellum, substantia nigra, hippocampus and fetal brain.
CC -1- DISEASE: Defects in CSDM1 may be a cause of oral and oropharyngeal
CC squamous cell carcinomas (OSCCs). Ref.5 and Ref.6 are however in
CC disagreement: while Ref.6 considers CSDM1 as a strong candidate
CC for OSCCs, Ref.5 thinks it is not.
CC -1- SIMILARITY: Belongs to the CSDM family.
CC -1- SIMILARITY: Contains 14 CUB domain.
CC -1- SIMILARITY: Contains 28 Sushi (CCP/SCR) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF333704; AAK73475.2; -;
DR EMBL; AY017307; AAG52948.1; -;
DR EMBL; AK126936; BAC86754.1; ALT_INIT.
DR EMBL; AY358174; AA088541.1; ALT_INIT.
DR EMBL; AB067477; BAB67783.1; -;
DR Genew; HGNC:14026; CSDM1.
DR MIM; 608397; -;
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PR00431; CUB; 6.
DR Pfam; PF00084; Sushi; 6.
DR PROSITE; PS01180; CUB; 14.
DR PROSITE; PS50923; SUSHI; 28.
KM Alternative splicing; Repeat; Signal; Sushi; Transmembrane.
FT SIGNAL 1 26 Potential.
FT CHAIN 27 3565 CUB and sushi multiple domains protein 1.
FT DOMAIN 27 3488 Extracellular (Potential).
FT TRANSMEM 3489 3509 Cytoplasmic (Potential).
FT DOMAIN 3510 3565 Potential.
FT DOMAIN 32 140 CUB 1.
FT DOMAIN 143 204 Sushi 1.

FT DOMAIN 208 312 CUB 2.
FT DOMAIN 347 409 Sushi 2.
FT DOMAIN 412 523 CUB 3.
FT DOMAIN 526 583 Sushi 3.
FT DOMAIN 585 693 CUB 4.
FT DOMAIN 696 757 Sushi 4.
FT DOMAIN 759 867 CUB 5.
FT DOMAIN 872 929 Sushi 5.
FT DOMAIN 931 1041 CUB 6.
FT DOMAIN 1044 1103 Sushi 6.
FT DOMAIN 1105 1213 CUB 7.
FT DOMAIN 1216 1276 Sushi 7.
FT DOMAIN 1278 1387 CUB 8.
FT DOMAIN 1390 1450 Sushi 8.
FT DOMAIN 1452 1560 CUB 9.
FT DOMAIN 1563 1624 Sushi 9.
FT DOMAIN 1626 1734 CUB 10.
FT DOMAIN 1740 1801 Sushi 10.
FT DOMAIN 1803 1911 CUB 11.
FT DOMAIN 1914 1973 Sushi 11.
FT DOMAIN 1975 2083 CUB 12.
FT DOMAIN 2086 2145 Sushi 12.
FT DOMAIN 2147 2258 CUB 13.
FT DOMAIN 2257 2318 Sushi 13.
FT DOMAIN 2320 2431 CUB 14.
FT DOMAIN 2431 2493 Sushi 14.
FT DOMAIN 2494 2555 Sushi 15.
FT DOMAIN 2556 2620 Sushi 16.
FT DOMAIN 2621 2678 Sushi 17.
FT DOMAIN 2679 2736 Sushi 18.
FT DOMAIN 2737 2794 Sushi 19.
FT DOMAIN 2795 2857 Sushi 20.
FT DOMAIN 2858 2915 Sushi 21.
FT DOMAIN 2919 2976 Sushi 22.
FT DOMAIN 2977 3035 Sushi 23.
FT DOMAIN 3036 3095 Sushi 24.
FT DOMAIN 3096 3153 Sushi 25.
FT DOMAIN 3154 3211 Sushi 26.
FT DOMAIN 3215 3273 Sushi 27.
FT DOMAIN 3274 3333 Sushi 28.
FT CARBOHYD 40 40 N-linked (GlcNAc...)
FT CARBOHYD 57 57 N-linked (GlcNAc...)
FT CARBOHYD 588 588 N-linked (GlcNAc...)
FT CARBOHYD 687 687 N-linked (GlcNAc...)
FT CARBOHYD 956 956 N-linked (GlcNAc...)
FT CARBOHYD 1016 1016 N-linked (GlcNAc...)
FT CARBOHYD 1035 1035 N-linked (GlcNAc...)
FT CARBOHYD 1185 1185 N-linked (GlcNAc...)
FT CARBOHYD 1198 1198 N-linked (GlcNAc...)
FT CARBOHYD 1400 1400 N-linked (GlcNAc...)
FT CARBOHYD 1455 1455 N-linked (GlcNAc...)
FT CARBOHYD 1573 1573 N-linked (GlcNAc...)
FT CARBOHYD 1645 1645 N-linked (GlcNAc...)
Query Match 67.0%; Score 11372.5; DB 1; Length 3565;
Best Local Similarity 64.1%; Pred. No. 0;
Matches 1970; Conservative 461; Mismatches 584; Indels 59; Gaps 9;
QY 1 MAGAPPALP-----LPCSLISD-----CCASQNRHSGVGPBELYKKQIELKSGV 47
DB 264 ISGTEAPSIWLTGMNLPSPVITSSKNWLRHFTSDSHRRK-GNAQGVKKALIELKSGV 322
QY 48 KLMPSKDNQKTVLVTVQVYSQGHMNCPPDGITFERGKRLGSDP-RIGSSVQFTCNBGYL 106
DB 323 KMLPSKDGSHKNSVLQSGVALVSHMCLPGIPIENGRRAQSDPSRVGAVQVFCEDNYVL 382
QY 107 QGSKRTICMKVSMFPAWSDHPVRCARBMCDALRGPSGITSPNPPIOVNNAHCYWI 166
DB 383 QGSKSITICQVTVTLTAMSDHPICARITGSLRGPSTVITSPNIPVQEDNAHCYWI 442
QY 167 TALNPSKVIKLAPEERDLERGYDTLVGDGQGDQDQKTVLYMSQNAQSPHPTGSRIP 226
DB 443 TTTPDPKVIKLAPEERDLERGYDTLVGDGAKGVDRSVLYV-----LTGSSVDP 492

QY 227 ---SMSGDIWROKWTVLEICRDISSSDARSQSVKSPKTSNAVELVAG-----TEHOG 278
 Db 433 LIVSNQNMWLA-----LQSDG-----IGSGFAVYVQOEIEKG 526
 QY 279 SCGDGIPAYGRREGSRPHHGDITLKECOPAFELVQKAITCOKNNQSAKPGCVSPSC 338
 Db 527 GCGDGPIDPAYGKRTSSFLHGDITLFECPAPAFELVGERVITCQONQMSGKRECVSFCF 586
 QY 339 FNFPSGCVULSPNPNPEYGNHLHCWMLILARPESRHLAINDIDVEPODFELVIXDQAT 398
 Db 567 FNFITASGILISPNNPEEYGNMNCWMLIIEPGRHILIFNDPDVEPODFELVIXDQGI 646
 QY 399 AEAPLYGTFSGNQULPSSITSSGHAVALREFOTDHSQKXGPNITFTTERHNCBPPGVPVN 458
 Db 647 SDITVLTGTFSGNEVPSCGLASSGHIVLEFQSDHSTTGKGFNITVTTGQNECHDPGIPIN 706
 QY 459 GKRFDLSIQLGSSISFLCDEGFLGTQSGEITTCVLKESVVMNSAVLRCEAPCGHILTSP 518
 Db 707 GRRGDRFLGSSVSFHCDDGFVKTQSGESITCIIQGNVWMSITVRCAPCGHILTAS 766
 QY 519 SGTLSPGMPGFYDALSCAWIEAORGYPIKITFDFKTEVNTDILEVFDGRTYSAPLI 578
 Db 767 SGVILPROMPGYVYKOSLHCWMIIEAKPGHSIKMTFDRFOTEVNTDILEVRDGPASSSPLI 826
 QY 579 GVHGTQVPOPLISTSNVLYLFFSTQKSHSDIGFOLRYETITLQSDHCLDPGI PVNGQRH 638
 Db 827 GEYHGTQAPQLISTGNFMWILFTTDSRSISIGLHIBESVTLFESBCLDPGIPVNGHRH 886
 QY 639 GNDVYVALYTFSCDSGYTLSDGPFLCEPFPWMSRALPSCCALCGGFIQSGSGTILSPG 698
 Db 887 GGDGIGISTVTFSCDPPYTLSDDEPLVCERNQNMNHALPSCDALCGGIQKSGTIVLSPG 946
 QY 699 FPDFYPPNNLNCWTMIETSHGKGVFTFTHLSGSHYLLITNGSGTOQLRDLTSSRLP 758
 Db 947 FPDFYPPNNLNCWTMIETSHGKGVFTFTHLSGSHYLLITNGSGTOQLRDLTSSRLP 758
 QY 759 APIISAGLYGNFTAVRPISDFSMSYEGFNITFSEYDLPECEPEVPAVYSIRKQLQGVGD 818
 Db 1007 HTIKAGFGFNFTQOLRFISDFSISYEGFNITFSEYDLPECEPEVPAVYSIRKQLQGVGD 1066
 QY 819 TLFPSCPGYRLGSLAINTLGGRRRLMSPLRCAVACNSYTGQGTILSPNPNVN 878
 Db 1067 SLTFSCGLYRLGSLAINTLGGRRRLMSPLRCAVACNSYTGQGTILSPNPNVN 878
 QY 879 NNHECIYSIQTOPKGIQOLKARAFELSEGDLKVYDGNMNSARLLGVFSHEMVGTYLNS 938
 Db 1127 NNHECIYSIQTOPKGIQOLKARAFELSEGDLKVYDGNMNSARLLGVFSHEMVGTYLNS 938
 QY 939 TSSSLWIDFTTDAENTSKGFEHLFSSPELICEDEPCTPKRGYKVHDEGHPAGSSVSFSCD 998
 Db 1187 TSSSLWIDFTTDAENTSKGFEHLFSSPELICEDEPCTPKRGYKVHDEGHPAGSSVSFSCD 998
 QY 999 PGYSLGSEBELCLSEGERRTWDRPLPTCAVACGCTVARGESQVLSPGIPAYEHNANCI 1058
 Db 1247 PGYSLGSEBELCLSEGERRTWDRPLPTCAVACGCTVARGESQVLSPGIPAYEHNANCI 1058
 QY 1059 WTEIABAGCTIGHLVPDTEEVHDLRIWDPVESGVLLKELSGPALPKDHLSTFNSV 1118
 Db 1307 WTEIABAGCTIGHLVPDTEEVHDLRIWDPVESGVLLKELSGPALPKDHLSTFNSV 1118
 QY 1119 LQSDPFTSKOGFALQFVSSTRTSNDGIRPONGSRSDSWEADSTVFCQDPCYALQ 1178
 Db 1367 LQSDPFTSKOGFALQFVSSTRTSNDGIRPONGSRSDSWEADSTVFCQDPCYALQ 1178
 QY 1179 SAEISCVKIEENFFWOPSPPTCIAPCGDLTGPSGUILSPNPEYPPGKECDMVKVYSP 1238
 Db 1427 SAEISCVKIEENFFWOPSPPTCIAPCGDLTGPSGUILSPNPEYPPGKECDMVKVYSP 1238
 QY 1239 DVIYALVNIENLEPGYDLFIATYDGDLSPLIGSFYGSOLPGRITSSNSLFLAFRSDA 1298
 Db 1487 DVIYALVNIENLEPGYDLFIATYDGDLSPLIGSFYGSOLPGRITSSNSLFLAFRSDA 1298

QY 1299 SVSNAQFVIDYTEMRESCFDPGSIKNGTRVGSDDLKLGSSVITYYCHGYEVEGTSTLSCI 1358
 Db 1547 SVSNAQFVIDYTEMRESCFDPGSIKNGTRVGSDDLKLGSSVITYYCHGYEVEGTSTLSCI 1358
 QY 1359 LGPDGKPVWNNRPVCTATACGGQYVSDGVULSPNPNQYTSQICILYVTVPKDYVVF 1418
 Db 1607 LGPDGKPVWNNRPVCTATACGGQYVSDGVULSPNPNQYTSQICILYVTVPKDYVVF 1418
 QY 1419 QPAFFHTALNDVVENVDGHSQSRLLSLSGSHGTSJPLATSNQVLIKFSKGLAPANG 1478
 Db 1667 QPAFFHTALNDVVENVDGHSQSRLLSLSGSHGTSJPLATSNQVLIKFSKGLAPANG 1478
 QY 1479 FHFVYQAVVRTSAITQSSVPEFRYGRIGSDFSVGAIVAREFECNSGVALQGSDEITCLPVP 1538
 Db 1727 FHFVYQAVVRTSAITQSSVPEFRYGRIGSDFSVGAIVAREFECNSGVALQGSDEITCLPVP 1538
 QY 1539 GALAONNVAAPTQVPCCGNLTERRGTLISPCFPERPLNSLNCWKIYVBEAGIQIOVY 1598
 Db 1787 GALAONNVAAPTQVPCCGNLTERRGTLISPCFPERPLNSLNCWKIYVBEAGIQIOVY 1598
 QY 1599 SEVTEQNDLSLEVPDADNTVTMLGSPGTTVPALINSTNOVLYHFYSDISVSAAGFHL 1658
 Db 1847 SEVTEQNDLSLEVPDADNTVTMLGSPGTTVPALINSTNOVLYHFYSDISVSAAGFHL 1658
 QY 1659 EYKTYGLSCPEPAPVPSNGVKTGERIYLVNDVVSFOCEPGVALQGHAIISCMPTVRWNY 1718
 Db 1907 EYKTYGLSCPEPAPVPSNGVKTGERIYLVNDVVSFOCEPGVALQGHAIISCMPTVRWNY 1718
 QY 1719 PPLICIAOCGGTVEEMEGEYILSPGPGYVPSNMDQSMKIALPVGCAHIOLENFTEBNH 1778
 Db 1967 PPLICIAOCGGTVEEMEGEYILSPGPGYVPSNMDQSMKIALPVGCAHIOLENFTEBNH 1778
 QY 1779 DYIEIRNRPYETSRMGRFSSSELPSLSTSHETTYVFNHSDSONRPGFLVQVAYELQ 1838
 Db 2027 DYIEIRNRPYETSRMGRFSSSELPSLSTSHETTYVFNHSDSONRPGFLVQVAYELQ 1838
 QY 1839 ECPDEPEPANGIVAGAGVNGQSVTEBCLPGYOLTGHVLTLCQNGTRNWMHPLPKCEVP 1898
 Db 2087 ECPDEPEPANGIVAGAGVNGQSVTEBCLPGYOLTGHVLTLCQNGTRNWMHPLPKCEVP 1898
 QY 1899 CGGNITSSNGTVSPSGPSPYSSQDCWMLITVPIGHCVRNLNLTLOTEPBGDFITIMDG 1958
 Db 2147 CGGNITSSNGTVSPSGPSPYSSQDCWMLITVPIGHCVRNLNLTLOTEPBGDFITIMDG 1958
 QY 1959 POQAPRLGIVTRSMAKTVOSSNOVYLKFRHDAAGIIPALAFSAVPLTKCPPTILP 2018
 Db 2207 POQAPRLGIVTRSMAKTVOSSNOVYLKFRHDAAGIIPALAFSAVPLTKCPPTILP 2018
 QY 2019 NAEVTEBERFNIQDIYRRCPLPGFTLVGNEILLCKXIGTYLOFEGRPPIGVHCFTEBL 2078
 Db 2267 NAEVTEBERFNIQDIYRRCPLPGFTLVGNEILLCKXIGTYLOFEGRPPIGVHCFTEBL 2078
 QY 2079 TDSGTGVLISQSPGSPYPOFOTCSWLVRBPDYNISLTVFELSEKQDEFEIPDPGSGQS 2138
 Db 2327 TDSGTGVLISQSPGSPYPOFOTCSWLVRBPDYNISLTVFELSEKQDEFEIPDPGSGQS 2138
 QY 2139 PLKALSGNSAPRIYVSSNSVYLRWSSDHAKNRKFKIRYAPYCSLPRAPLHGFTLG 2198
 Db 2387 PLKALSGNSAPRIYVSSNSVYLRWSSDHAKNRKFKIRYAPYCSLPRAPLHGFTLG 2198
 QY 2199 QTSQTOPGSIHFGCNAQYRLVGHSMALCTRHPOGYHLSMAIPLCOALSGLBEAPANGM 2258
 Db 2447 QTSQTOPGSIHFGCNAQYRLVGHSMALCTRHPOGYHLSMAIPLCOALSGLBEAPANGM 2258
 QY 2259 VFGKEVYVTKAVYSGSEGYHQAABATACLDITGLMSNRNVPPOCVPTCPDVSISIV 2318
 Db 2507 VFGKEVYVTKAVYSGSEGYHQAABATACLDITGLMSNRNVPPOCVPTCPDVSISIV 2318
 QY 2319 EHGKRLIIFETQYQFOQALMLICDPGYTYTGQVIVICQANGKSLSDSPITCIIISGEL 2378
 Db 2567 EHGKRLIIFETQYQFOQALMLICDPGYTYTGQVIVICQANGKSLSDSPITCIIISGEL 2378
 QY 2379 PIPNGHRIQTLISVYATAIFSCNSGTYLVGSRVRECMANGLMSGSEVCLACHGCTPBP 2438

Db	2627	SPPPKNNKIGTLTYGAATATFTCNNGTTLVSHSHRECLANGLMSSSEFRCLAGHCGSDP	2686
Qy	2439	IVNGHINGENYSYRGSVVYQCNAGFRLLIGMSVRIQQDDHHSKTPFCVPVPTTCGPGNPV	24398
Db	2687	IVNGHISGDGSESYRDTVVYQCNPGFRLVGTSTVRIQLADHKMSGGCPVCPVPTTCGPGNPA	2746
Qy	2499	NGLIQGNPNLNDVYKFCVNCBGMAMEGAARQQCLASGQWSDMLPFCRIILNCTDGHQNS	2558
Db	2747	HGFNTGSSSFNLNDVNFCTNCTGYLLQGVSRACQKRNNGWSSPLPFCRVNCSDEPVENA	2806
Qy	2559	VROVYASGPHRPFSGFTVSYRCSNHGFYLLGTPVLSQGDQGMWDRRPPQCLVSCGHPSP	2618
Db	2807	IRHQGNPFPSPEFYEMSLIHYCKKGFFHLGSSALTCMANGLMDSLPKCLALISGHEVP	2866
Qy	2619	PHSQWSDSYTVGAAVVRYSICIGKRTLNVNSITRMCGLDGHWMTGSLPHCSGTSVVGCDPGI	2678
Db	2867	ANAVLTGELFTYGAAVHYSCRGSESLIGNDRFVQCEDSHWSGALPHCTGNPPGCGDPT	2926
Qy	2679	PAHGRLDSDPDPCGVMAFPSCBAGHYVLRGSSERTQANGSNSSGQPECGVISCNPGTFS	2738
Db	2927	PAHGRLDDEFTYKSLFRSCEMGHQLGSPERTCLLNGSWSGLOPVEAASCNPGTPT	2986
Qy	2729	NARVVSQGLVFPSSIVYECRGEYVATGLSHRCSHGVTMGVSDPECLVINCSPGPIAN	2798
Db	2967	NGMIVSSGGLFSSSVIYACMEGYTSGGLMTRHCTANSTMTGTADCTIISCGDGTLAN	3046
Qy	2799	GLRLGNDRPNKTVYQCPVPGYMSHRVSVLSCTKDRTNMGTKPVCKALKCKPPLIPN	2858
Db	3047	GIQGFDTPTPKNTSYQCNPGVMEAVYASATIRCKDRMNPSPKVCYCAVLCPQDPYQN	3106
Qy	2859	GKVVSGDTMWGSSVTVYACLQGYQLSPVFTCEGNGSVTGLPQCFPVFCGDPPVPSRGR	2918
Db	3107	GTVESSDTRWGSISYSQMDGQYQLSHSAILSCBGRGVWKGPCLPQCFVCGDPPGIPAGR	3166
Qy	2919	REDGFYSRSSVSFCHPPLVYVGSRRFCQSDGWSGTQSPSCIDPILTTCADPGVQFG	2978
Db	3167	LSGKSFYTKSEVFPFOCKSPFLVGVSSRRVCCADGWSGIQPTCIDPAHNTPDPTPTPG	3226
Qy	2979	IQNNSQGYQVGSVTVFCRCQGYLLQGSYTRTCLPLVMTSGTPPDCVPHHCQRPETPTAN	3038
Db	3227	IQNSRGYEGVGSVTFVFRCKGHIOGSIITRTCLANLVTSGIQTECIPHACQRPETPAHD	3286
Qy	3039	VGALDLPMSGYTLI 3052	
Db	3287	VRAIDLPTFGYTLV 3300	

RESULT 3

CSML_MOUSE	STANDARD;	PRT;	3564 AA.
ID	CSML_MOUSE		
AC	Q92JL3; Q8BUV1; Q8BYO3;		
DT	29-MAR-2004 (Rel. 43, Created)		
DT	29-MAR-2004 (Rel. 43, Last sequence update)		
DT	05-JUL-2004 (Rel. 44, Last annotation update)		
DE	CUB and BUSH1 multiple domains protein 1 precursor.		
GN	Name=Cemdi;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	STRAIN=C57BL/6;		
RX	MEDLINE=21365705; PubMed=11472063; DOI=10.1006/geno.2001.6587;		
RA	Sun P.C., Uppeluri R., Schmidt A.P., Pashia M.E., Quant E.C.,		
RA	Sunwoo J.B., Gollin S.W., Scholnick S.B.;		
RT	"Transgenic map of the 8p23 putative tumor suppressor region.";		
RL	Genomics 75:17-25(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 3250-3564 FROM N.A.		
RP	(ISOFORM 2).		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum, and Hypothalamus;		

RE	MEBLIN=22351683; PubMed=12466851; DOI=10.1038/nature01266;
RA	Nkaido I. Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
RA	Baldarelli R., Hill D. P., Bult C., Hume D. A., Quackenbush J.,
RA	Schmitt L. M., Knapin A., Matsuda H., Batalov S., Beisel K. W.,
RA	Blake J. A., Brad D., Bruneir V., Chochia C., Corbani L. E., Cousins S.,
RA	Dalla E., Drganti T. A., Fletcher C. F., Forrest A., Frazer K. S.,
RA	Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA	Grimmond S., Guestinich S., Hirokawa N., Jackson I. J., Jarvis E. D.,
RA	Kanai A., Kawaji H., Kawasawa Y., Kedzierski R. M., King B. A.,
RA	Konagaya A., Kurochkin I. V., Lee Y., Lemhard B., Lyons P. L.,
RA	Maglott D. R., Maltais L., Marchionni L., McKenzie L. T., Miki H.,
RA	Nagashima T., Numata K., Okido T., Pavan W. J., Petrea G., Pesole G.,
RA	Petrovsky N., Pillai R., Pontius J. U., Qi D., Ramachandran S.,
RA	Ravasi T., Reed J. C., Reed D. J., Reid J., Ring B. Z., Ringwald M.,
RA	Sandelin A., Schneider C., Sempke C. A., Setou M., Shimada K.,
RA	Sultana R., Takenaka Y., Taylor M. S., Teasdale R. D., Tomita M.,
RA	Verardo R., Wagner L., Wahlestedt C., Wang Y., Watande Y., Wells C.,
RA	Wilmink L. G., Wyshnaw-Boris A., Yangisawa M., Yang I., Yang L.,
RA	Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA	Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA	Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA	Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA	Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA	Yasunishi A., Yoshino M., Waterston R., Lander E. S., Rogers J.,
RT	Bitney E., Hayaishi-Zeki Y.!
RT	"Analysis of the mouse transcriptome based on functional annotation of
RL	60,770 full-length cDNAs".
RL	Nature 420:563-573(2002).
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC	-1- ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=3;
CC	Name=1;
CC	IsoId=Q923L3-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId=Q923L3-2; Sequence=VSP_009037;
CC	Note=No experimental confirmation available;
CC	Name=3;
CC	IsoId=Q923L3-3; Sequence=VSP_009036;
CC	Note=No experimental confirmation available;
CC	-1- SIMILARITY: Belongs to the GMD family.
CC	-1- SIMILARITY: Contains 14 CUB domains.
CC	-1- SIMILARITY: Contains 28 Sushi (CCP/SCR) domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AY017475; AAC54083.1; -
DR	EMBL; AK038679; BAC30095.1; ALT_INIT.
DR	EMBL; AK082377; BAC38482.1; -
DR	HSSP; Q9UCV4; INZI.
DR	MGD; MGI:2137383; Camd1.
DR	InterPro; IPR000859; CUB.
DR	InterPro; IPR000436; Sushi_SCR_CCP.
DR	Pfam; PF00431; CUB; 14.
DR	Pfam; PF00084; Sushi; 28.
DR	PROSITE; PS01180; CUB; 14.
DR	PROSITE; PS50923; SUSHI; 26.
KW	Alternative splicing; Repeat; Signal; Sushi; Transmembrane.
FT	SIGNAL
FT	1
FT	29
FT	POTENTIAL
FT	30
FT	3564
FT	CHAIN
FT	30
FT	3487
FT	CUB and sushi multiple domains protein 1.
FT	TRANSMEM
FT	3488
FT	3508
FT	POTENTIAL
FT	DOMAIN
FT	3509
FT	3564
FT	Cytoplasmic (Potential).
FT	DOMAIN
FT	32
FT	140
FT	CUB 1.
FT	DOMAIN
FT	143
FT	204
FT	Sushi 1.
FT	DOMAIN
FT	208
FT	312
FT	CUB 2

FT	DOMAIN	347	408	Sushi 2.
FT	DOMAIN	411	522	CUB 3.
FT	DOMAIN	525	582	Sushi 3.
FT	DOMAIN	584	692	CUB 4.
FT	DOMAIN	695	756	Sushi 4.
FT	DOMAIN	758	866	CUB 5.
FT	DOMAIN	871	928	Sushi 5.
FT	DOMAIN	930	1040	CUB 6.
FT	DOMAIN	1043	1102	Sushi 6.
FT	DOMAIN	1104	1212	CUB 7.
FT	DOMAIN	1215	1275	Sushi 7.
FT	DOMAIN	1277	1386	CUB 8.
FT	DOMAIN	1389	1449	Sushi 8.
FT	DOMAIN	1451	1559	CUB 9.
FT	DOMAIN	1562	1623	Sushi 9.
FT	DOMAIN	1625	1733	CUB 10.
FT	DOMAIN	1739	1800	Sushi 10.
FT	DOMAIN	1802	1910	CUB 11.
FT	DOMAIN	1913	1972	Sushi 11.
FT	DOMAIN	1974	2082	CUB 12.
FT	DOMAIN	2085	2144	Sushi 12.
FT	DOMAIN	2146	2257	CUB 13.
FT	DOMAIN	2256	2317	Sushi 13.
FT	DOMAIN	2319	2430	CUB 14.
FT	DOMAIN	2430	2492	Sushi 14.
FT	DOMAIN	2493	2554	Sushi 15.
FT	DOMAIN	2555	2619	Sushi 16.
FT	DOMAIN	2620	2673	Sushi 17.
FT	DOMAIN	2678	2735	Sushi 18.
FT	DOMAIN	2736	2793	Sushi 19.
FT	DOMAIN	2794	2856	Sushi 20.
FT	DOMAIN	2857	2914	Sushi 21.
FT	DOMAIN	2918	2975	Sushi 22.
FT	DOMAIN	2976	3034	Sushi 23.
FT	DOMAIN	3035	3094	Sushi 24.
FT	DOMAIN	3095	3152	Sushi 25.
FT	DOMAIN	3153	3210	Sushi 26.
FT	DOMAIN	3214	3272	Sushi 27.
FT	DOMAIN	3273	3332	Sushi 28.
FT	CARBOHYD	40	40	N-linked (GlcNAc. . .)
FT	CARBOHYD	57	57	N-linked (GlcNAc. . .)
FT	CARBOHYD	587	587	N-linked (GlcNAc. . .)
FT	CARBOHYD	686	686	N-linked (GlcNAc. . .)
FT	CARBOHYD	955	955	N-linked (GlcNAc. . .)
FT	CARBOHYD	1015	1015	N-linked (GlcNAc. . .)
FT	CARBOHYD	1034	1034	N-linked (GlcNAc. . .)
FT	CARBOHYD	1184	1184	N-linked (GlcNAc. . .)
FT	CARBOHYD	1197	1197	N-linked (GlcNAc. . .)
FT	CARBOHYD	1399	1399	N-linked (GlcNAc. . .)
FT	CARBOHYD	1454	1454	N-linked (GlcNAc. . .)
FT	CARBOHYD	1572	1572	N-linked (GlcNAc. . .)
FT	CARBOHYD	1644	1644	N-linked (GlcNAc. . .)
FT	CARBOHYD	1792	1792	N-linked (GlcNAc. . .)
FT	CARBOHYD	1805	1805	N-linked (GlcNAc. . .)
FT	CARBOHYD	1882	1882	N-linked (GlcNAc. . .)
FT	CARBOHYD	2018	2018	N-linked (GlcNAc. . .)
FT	CARBOHYD	2149	2149	N-linked (GlcNAc. . .)
FT	CARBOHYD	2154	2154	N-linked (GlcNAc. . .)
FT	CARBOHYD	2187	2187	N-linked (GlcNAc. . .)
FT	CARBOHYD	2358	2358	N-linked (GlcNAc. . .)
FT	CARBOHYD	2394	2394	N-linked (GlcNAc. . .)
FT	CARBOHYD	2400	2400	N-linked (GlcNAc. . .)
FT	CARBOHYD	2445	2445	N-linked (GlcNAc. . .)
FT	CARBOHYD	2470	2470	N-linked (GlcNAc. . .)
FT	CARBOHYD	2503	2503	N-linked (GlcNAc. . .)
FT	CARBOHYD	2605	2605	N-linked (GlcNAc. . .)
FT	CARBOHYD	2750	2750	N-linked (GlcNAc. . .)
FT	CARBOHYD	2761	2761	N-linked (GlcNAc. . .)
FT	CARBOHYD	2795	2795	N-linked (GlcNAc. . .)
FT	CARBOHYD	2894	2894	N-linked (GlcNAc. . .)
FT	CARBOHYD	2963	2963	N-linked (GlcNAc. . .)
FT	CARBOHYD	3022	3022	N-linked (GlcNAc. . .)
FT	CARBOHYD	3056	3056	N-linked (GlcNAc. . .)

```
Db 1067 LAFPCGQYRLEGATKULTCGGGRVMSAPLPRCAEGCASVKNEBGLTSLPNFESHYN 1126
Qy 880 NHECIYISIQTPGKGIOLKAPAFELSEBDVLKATDGNNSAKLLGVSHSMEGVTLANST 939
Db 1127 NHECIYKIEETAGKIHRLARTFQLFEGDTLKVYDGKSSRSRGVFRSFMGLVJNST 1186
Qy 940 SSSMLPBITDAENTSKGFLHFSFELIKCEDPCTPKFGYKHYDEGHFASVSFSCDP 999
Db 1187 SNYLRLEPNYNGSTOAGFOLTYTSFDLVKCEDPFIYNYGTRIKDDGHFTDTVLYSCNP 1246
Qy 1000 GYSLRGESEELCLSGERRTMDRPLPTCAECGTVRGEVSGOVLSPGYAPAEHNLNCIW 1059
Db 1247 GYAHGSSSTLTCLSGDRRWMDKPMPSVACGGLVHAATSGRILSLPGYAPAYDNHHCWTW 1306
Qy 1060 TIEBAGCTTGLHFLVPTDEEVHVLKWDGVRSGVULKELSGPALPKDLHSTFNSVYL 1119
Db 1307 TIEADPKTISLHFLVPTDETHADILKWDGVPVDSNILLKEMSGSALPEDIHSTFNSLTL 1366
Qy 1120 QFSTDPFTSKGFAIOFSVSTATSCNDPGRIPONGSRSGDSWEAGDSTVFOCDPGYALQGS 1179
Db 1367 QFSDPFTSKGFSIOFSTSLASTCNDPQMONGRIGDSREPGDTTTFQCDPGYALQGP 1426
Qy 1180 AELSCVAKIENRFWQPSPTICIAFCGGDLTGPSSGVILSPNYBEPPYKCEKDCMAYTVSPD 1239
Db 1427 AKICVQNLNRRFFMQPDPSPSCIAACGGNLTPAGVILSPNYPOYPRGKECDMRIXNPD 1486
Qy 1240 YVILAVNIFNLBEGYDFLHYDGRDSLPLISGFYSGQLPGRIESSNSLFLAFRSDAS 1299
Db 1487 FVILILFKFSFMSBPSYDFLHYEBGEDSNSPLIGFQCSQAPERIESSGNSLFLAFRSDAS 1546
Qy 1300 VSNAGFIIDYENRESQFDPGSIKNGTRVGSDLKLSVTVYHGGYEVGCTSTLSCIL 1359
Db 1547 VGLSGFPIEFKEKERACFPDGNIMNGRIGTDPRKLSVTYQCDSGKTIIDPSISICVT 1606
Qy 1360 GPDGKPYMNNRPVCTAPCCGQYVSGDVVLSPNYPQNYTSGQICLYFVTVPKDYVVFQ 1419
Db 1607 GADGKPSMDRALPACQAPCCGQYVSGDVVLSPNYPHNYTAGQMCVYSITVYKXFFVFGQ 1666
Qy 1420 PAFHTYLANDVVEYVHDSHSHSLSLSGSHTEBSLPLATSNQVILKFSAGKLAAPRGF 1479
Db 1667 FAYFOTILNDLAEPLFDGTHPOARLLSLSGSHSEGTPLATSNQVILKFSAGSASARGF 1726
Qy 1480 HFVYQAVPRTSATQCSSVPEBRVGRKLGSDFSVGAIVRECNSSVLAQSGSEIECLVPG 1559
Db 1727 HFVYQAVPRTSDQCSSVPEBRVGRKIGSEFSAGSIVRPECNPGYLLQGSTALRCQSVPN 1786
Qy 1540 ALAQMNVSAPTCVVPCCGNLTERGTTILSPGFPEPYLNSLNCWKIIVBEGAGIQIOVVS 1599
Db 1787 ALAQMNVTIISCVVPCSGNFTQGRGTTILSPGYPERPYGNLNCWKIIVSEBSGIIQIOVIS 1846
Qy 1600 FVTEQWMSLEVPDADNVTYMLGSPSGTVPALLNSTSNQVILHAFYSDISVSAAFGHLE 1659
Db 1847 FATEQWMSLEIHDGDMTAPRLGSPGTVPALNSTSNQVILHAFSDISVAAFGHLE 1906
Qy 1660 YKTVGLSSCEBPAPRPSGVYKGERVLYNDVVSFOCEGVALOGHAHISCMHGTTRMMYR 1719
Db 1907 YKTVGLAACOBEPALPSNGIKGDRYVNDVLSFQCEPGYTLQGSNHSICMGTVRRMNYR 1966
Qy 1720 PPLCIAOCGGTVEBEGVILSPGFPGNVPNSMDCSMKIALPVGGAHIOFLNFSTERNHD 1779
Db 1967 SPCLIAOCGGTILMSGVILSPGFPGSVPNLDCWKIISLPYIGGAHIOFLNFSTERNHD 2026
Qy 1780 YIEIRNCPYETSRMGRFSGSELPSSLSLSTSHETTVYFHSDSHONRPGFKLEYAYELQE 1839
Db 2027 YLIEIONPYPHSSPPMGOFSGPDLPSTLSLSTHETTLIRFYSDHSQNRGQFKLSYQAYELQN 2086
Qy 1840 CPDPEBPANGIVRAGVNGQSVTFECLPGYQLTGHVVLTCQGHSTNNMHPKCEVPC 1899
Db 2087 CPDPEBPANGFMINDSDSVSGSISFECYPGYILGHVVLTCQGHSTNNMHPFERCACP 2146
Qy 1900 GGNITSSNGTVYSPGFPSPYSSQDCVWLITVPIGHVRLNLSLQTEPSGDPFTINDGP 1959
Db 2147 GYNTVSONGTIVSPGFDEXPILKDCMLVTVPPGHGVYINFLLQTEAVNDYIAVMDGP 2206
Qy 1960 QOTAPRLGVTRSAKTKVSSNOVULLKTHRDAAATGCIPIALASAPLTKCPPTILPN 2019
Db 2207 DQNSPOLGVSPGNAPETAAVSTNQVULLKTHSDPSNMGFFVLANHAOLKRCPPPAVPO 2266
Qy 2020 AEVVTENBEPNIGIVAYRCILPGFTLVGNEILTKLTGVYLOFEGEPPICEVHCCTNELLT 2079
Db 2267 ADLITEBDEBEIDFVYKQCHPGFTLLGSDTLTKLSSQLLPQSPPTCEACQCANEVRT 2326
Qy 2080 DSTGVILSQSPGSPYQFOTCSMLRVPEPDYVNSLTYEYFLSEKQYDFEIPDQSPGSP 2139
Db 2327 ESSGVILSPGPGNVPFNSQICAMGIIKYPMPNITLFDVTFQSEKQFALAEVFDQSSGRSP 2386
Qy 2140 LKRLSGNTYSAPLIVTSSNSVYLRMSDHAHNRKGRKIRYSAPYCGLPAPALHGFLLQ 2199
Db 2387 LVLVLSNHTHEQSNFTSRSHMLYLRMSTDHATSKKGFKIRYAAVCYLSJTLRNGJILNK 2446
Qy 2200 TSTQPGSIHPGCGAGYRLVGHSMALCTRRPOGYHLMSEAIPLCQALSCGLPEAPKXGMV 2259
Db 2447 TAGAVSKAHYFCXPGYRMTGHSNATCRNRPVGYQYQDMSAPLCQAVSCGIIPEAPNGSF 2506
Qy 2260 FGEKYTVGTRAVYSCSEGYHLOAGAETAACDLTGMLSNBNVPEQCVVTCPPVSSISVE 2319
Db 2507 TGNEFTLDSKYTYECNEGFILDASQEAATVQCEBGLMSNKGKPTCKPVPQPSIEQLSE 2566
Qy 2320 HGRNRLIFETQYQFOALMLICDPGYTYTQGVYIRCOANGKWSIGDSTPTCRILISCEBLP 2379
Db 2567 HVMRLVSGSILNEGAVLLSCSPGYFLQCORLLOCCANGTWSFEEDRPPKCVISCSLS 2626
Qy 2380 IPPNGHRIGLTVYGAFAIPSCNSGYTLVGSRVRECMANGMISGEVRCLAGHGTPEPI 2439
Db 2627 FPPNGKIGITLITYGALPAITCMTGYTLVSHVRECLANGMISSETRCLAGHGSDDPI 2686
Qy 2440 VNGHINENYSYKSSVYQCNAGFRILGMSVRILCOQDHMSGKTPFCVPTTCGHPGPNV 2499
Db 2687 VNGHISDGRSYNDVTVYQCNPGFRILGTVSVRLICRTTSRGRRLTVCPVITTCGHPGPAH 2746
Qy 2500 GLTQGNOPNLNDVYKFCNPGNYMAEGARSOCLASGWSMDLPTCRILINCTDPGHQENS 2559
Db 2747 GLTNGTEFNINDVILNFCHTGYRLOGASRAQCSNMGWSPPLRCRVVNSDPSGVNNAV 2806
Qy 2560 ROYVASGPHRPFEGTTVYRCHNGFYLLGTPVLSQCGDGMWRDRPOCLVLSGHPSP 2619
Db 2807 RHQGNPESFEGTSMYHCKTGTFYLLGSSALTCMAAGLMDSLPRCLALISGHPVPA 2866
Qy 2620 HSGWSDSYTVGAVVRYSICGKRTLVGNSTRMCGLDGHTGSLPHCSGTSVGVCGDGP 2679
Db 2867 NAVLTGSLFTYGATVQYSCAGGQILTGNSTRVCOEDBSHMGSLPHCGSNRPGFCGDPTR 2926
Qy 2680 AHGIRLGSDFDPTWRFSCBAGHVLRGSEERTCOANGWSGQPECGVISCNPGTPSN 2739
Db 2927 AHGSRILDEKFTKSLIRFSCBEMGHOLRGFAERTCLVNGWSGQVPVCEAVSCNPGTPTN 2986
Qy 2740 ARVYFSDGIVFSSSIYVECEGYATGSLRHSVNGTWMGSPBECLVINGCPDGIANG 2799
Db 2987 GMLTSSDGLIFSSSVIYACMEGKTSGLMRHCTLANSTGTALDCTTISGPDGTLPNG 3046
Qy 2800 LRLANDPFRYKTYTVOCPVYMMESHNVSVLSTCKTRTNMGTPVCKALCKCPPLIPNG 2859
Db 3047 IQGTGDTFPMKTVSYQCNPNPXYLMBPRTSPILRTCKGTQWQNSRPLCAVALCQNPBPYPNG 3106
Qy 2860 KVVGSDFMGSVTVYACLBGTQLSLPAVFTCEGNGVTGELPOCFVYFCDDPGVPSRGR 2919
Db 3107 KVGSGDFRMWASISISYCVDSYOLSHSAILSCERGVKVGAVPOCLPVFCDDPGVPABGR 3166
Qy 2920 EDGFSYRSVSPSCHPRLVTVGSPPRFQCSQDGTWGSCTQSCIDPPLTTGADGVPOFGI 2979
Db 3167 SGKSFYKSSVFTQCKPFPVTVGSSSRKTCQADGIMSGIOPTCIDPAHTACPDGTPHFI 3226
Qy 2980 QNNSQGYOVGSYTVLFRCKQGYLLQGSTTRTCLPNLTWSGTPPCVPHHCQPEPTTANV 3039
Db 3227 QNNSKGYEVGSYTVFRCKQKYNHIOGSTTRTCLANLTWMSGIGTCECIPHACQPEPTANADV 3286
```

Qy 3040 GALDIPSMGYTLI 3052
 Db 3287 RAIDLPACGYTLV 3299

RESULT 4
 CSM3_HUMAN STANDARD; PRT; 3670 AA.
 ID_CSM3_HUMAN
 AC Q72407; Q96P23;
 DT 29-MAR-2004 (Rel. 43, Created)
 DT 29-MAR-2004 (Rel. 43, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DB CUB and sushi multiple domains precursor.
 GN Name=CSMD3; Synonyms=KIAA1894;
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RX MEDLINE=22788796; PubMed=12906867; DOI=10.1016/S0888-7543(03)00149-6;
 RA Lau M.L., Scholnick S.B.;
 RT "Identification of two new members of the CSMD gene family.";
 RL Genomics 82:412-415(2003).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), TISSUE SPECIFICITY, AND
 RP VARIANTS MET-182 AND HIS-3584.
 RC TISSUE=Brain, and Testis;
 RX PubMed=12943675; DOI=10.1016/S0006-291X(03)01555-9;
 RA Shimizu A., Asakawa S., Shimizu N.;
 RT "A novel giant gene CSMD3 encoding a protein with CUB and sushi
 RT multiple domains: a candidate gene for benign adult familial myoclonic
 RT epilepsy on human chromosome 8q23.3-q24.1.";
 RL Biochem. Biophys. Res. Commun. 309:143-154(2003).
 RN [3]
 RP SEQUENCE OF 624-3670 FROM N.A. (ISOFORM 4), AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=21456161; PubMed=11572484;
 RA Nagase T., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXI.
 RT The complete sequences of 60 new cDNA clones from brain which code for
 RT large proteins.";
 RL DNA Res. 8:179-187(2001).
 RN [4]
 RP SEQUENCE OF 2234-3670 FROM N.A. (ISOFORM 5).
 RC TISSUE=Liver;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahori K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoi T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Nishimura K., Iehibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hattori T.,
 RA Kusano J., Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togashi S., Komai F., Hara R., Takuchi K., Arita M.,
 RA Imose N., Matsuhiro K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Komiyama M., Tashiro H., Taniguchi A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=1;
 CC IsoId=Q72407-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q72407-2; Sequence=VSP_009047;
 CC Name=3;
 CC IsoId=Q72407-3; Sequence=VSP_009048, VSP_009049;
 CC Name=4;
 CC IsoId=Q72407-4; Sequence=VSP_009050;
 CC Note=No experimental confirmation available;
 CC Name=5;
 CC IsoId=Q72407-5; Sequence=VSP_009051, VSP_009052;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Weakly expressed in most tissues, except in
 CC brain. Expressed at intermediate level in brain, including
 CC cerebellum, substantia nigra, thalamus, spinal cord, hippocampus
 CC and fetal brain. Also expressed in testis.
 CC -1- SIMILARITY: Belongs to the CSMD family.
 CC -1- SIMILARITY: Contains 14 CUB domains.
 CC -1- SIMILARITY: Contains 28 Sushi (CCP/SCR) domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AY210419; AAC34702.1; -;
 DR EMBL; AB114604; BAC82443.1; ALT_INIT.
 DR EMBL; AB114605; BAC82444.1; -;
 DR EMBL; AB067481; BAB67787.2; -;
 DR EMBL; AK126252; BAC6505.1; ALT_INIT.
 DR Genew; HGNC:19291; CSMD3.
 DR MIM: 608399; -;
 DR Interpro; IPR000859; CUB.
 DR Interpro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00431; CUB; 14.
 DR Pfam; PF00084; Sushi; 27.
 DR ProSite; PS01180; CUB; 14.
 DR ProSite; PS0923; Sushi; 28.
 KW Alternative splicing; Polymorphism; Repeat; Signal; Sushi;
 KW Transmembrane.
 FT SIGNAL 1 23
 FT CHAIN 24 3670 Potential.
 FT DOMAIN 24 3593 CUB and sushi multiple domains protein 3.
 FT TRANSMEM 3594 3614 Extracellular (Potential).
 FT DOMAIN 3615 3670 Potential.
 FT DOMAIN 28 136 Cytoplasmic (Potential).
 FT DOMAIN 139 200 Sushi 1.
 FT DOMAIN 204 308 Sushi 2.
 FT DOMAIN 447 508 Sushi 2.
 FT DOMAIN 511 622 Sushi 3.
 FT DOMAIN 625 682 Sushi 3.
 FT DOMAIN 684 792 Sushi 4.
 FT DOMAIN 795 856 Sushi 4.
 FT DOMAIN 858 966 Sushi 5.
 FT DOMAIN 971 1028 Sushi 5.
 FT DOMAIN 1030 1140 Sushi 5.
 FT DOMAIN 1143 1202 Sushi 6.
 FT DOMAIN 1204 1312 Sushi 7.
 FT DOMAIN 1315 1375 Sushi 7.
 FT DOMAIN 1377 1486 Sushi 8.
 FT DOMAIN 1489 1549 Sushi 8.

```

FT DOMAIN 1551 1659 CUB 9.
FT DOMAIN 1662 1723 Sushl 9.
FT DOMAIN 1725 1833 CUB 10.
FT DOMAIN 1839 1900 Sushl 10.
FT DOMAIN 1902 2010 CUB 11.
FT DOMAIN 2013 2072 Sushl 11.
FT DOMAIN 2074 2182 CUB 12.
FT DOMAIN 2185 2244 Sushl 12.
FT DOMAIN 2246 2357 CUB 13.
FT DOMAIN 2356 2417 Sushl 13.
FT DOMAIN 2419 2530 CUB 14.
FT DOMAIN 2530 2592 Sushl 14.
FT DOMAIN 2593 2654 Sushl 15.
FT DOMAIN 2655 2719 Sushl 16.
FT DOMAIN 2720 2777 Sushl 17.
FT DOMAIN 2778 2835 Sushl 18.
FT DOMAIN 2836 2893 Sushl 19.
FT DOMAIN 2894 2955 Sushl 20.
FT DOMAIN 2956 3013 Sushl 21.
FT DOMAIN 3017 3074 Sushl 22.
FT DOMAIN 3075 3133 Sushl 23.
FT DOMAIN 3134 3193 Sushl 24.
FT DOMAIN 3194 3251 Sushl 25.
FT DOMAIN 3252 3309 Sushl 26.
FT DOMAIN 3313 3371 Sushl 27.
FT DOMAIN 3372 3431 Sushl 28.
FT CARBOHYD 36 53 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 53 53 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 324 324 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 372 372 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 687 687 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 786 786 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 929 929 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1055 1055 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1089 1089 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1134 1134 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1243 1243 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1499 1499 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1554 1554 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1572 1572 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1744 1744 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1892 1892 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 1982 1982 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2118 2118 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2249 2249 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2254 2254 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2287 2287 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2458 2458 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2500 2500 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2574 2574 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2674 2674 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2705 2705 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2825 2825 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2895 2895 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 2915 2915 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3062 3062 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3121 3121 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3130 3130 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3157 3157 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3171 3171 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3181 3181 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3239 3239 N-linked (G1cNAC. . .) (Potential).
FT CARBOHYD 3327 3327 N-linked (G1cNAC. . .) (Potential).

Query Match 62.4%; Score 10591.5; DB 1; Length 3670;
Best Local Similarity 60.1%; Pred. No. 0;
Matches 1816; Conservative 501; Mismatches 665; Indels 41; Gaps 6;

```

```

Db 471 VQFSCDEBDYVLQGAASITQCRIAEFAWSDHRRPVCKYKTCGSMVLQSGFTSPNFPFQ 530
156 YDNNAHCVWITITANPKVITKLAEEFEDLERGYTTLVVGCGGDDQDKTUYMQMNCSD 215
531 YDSNAQCWVITAVNTKVIQINEEFDELDGYTTLIGGEGVDPRVYQV----- 583
216 SPHTPGRIPE---SMGDIWRQKVTALICRDISSDARSGRKPKTSNAVELVAPG 272
584 ---LTGSFVDPDLVISMSSQWMLHQT-----DESQSVGFKNVK----- 620
273 TEIBQSCGDPGIPAYRRREGSRPHGDTLKFEQCPAFELVGOAKITCKNNQMSAKPG 332
621 -EIKESGDPGPTLYGIRREGDGSNRDVAREFCQGFELIGKSIYQENQMSANLPI 679
333 CVESCFNFTSPSGVULSPNRYEYGHNLICWILLARPESRITLANDIDVEQDPVLY 392
680 CIFPCLSNFTAPMGVTLSPDPYEGYNNLNCIWTIISDPSGRHLSPNDPDLFESQPFPLA 739
393 IKDQATAEAPVLTGFSGNOLPSISITSGHVARLEFQTDHSTGKSGFNITFTPHNDCPD 452
740 VKQDSSESPLTGFTGAEPVSHLTSNHLIRLEFQADHSMGSGFNITFTPHNDCPD 799
453 PGVFNKRRPDSLIQGLSSISFLCDEBPLGTQSGETITCVLKESGVYVNSAVLRCEAPCG 512
800 PGIRINARRGDNFQGLSSISVICBGFKITQGETITCILMDGKVMWSGLIPKCAPCG 859
513 GHLSBPGTILSPMPGPFYKDALSACAVITAEOPVPIKITPDRKTEVNTDLEVRGRT 572
860 GHFAPSGVILSPMPGYKDSLNCENVIEBPSHSIKITFERQTELVNDAVLEHVGPN 919
573 YSAPLIVYNGTQVPOPLISTSNLYLFFSTDKSHSDIGFQLRJETITLLOSDHCLDGP 632
920 LLSPLLSYNGTQVPOPLFSSSNFYLLFTTDSNRSNNGKIKIHESVTVNTSCLDGP 979
633 VNGQRHNDYVYALVTFSCDSGYTSLDGPBCEPNFOWSRALPSCALCGFIQSSG 692
980 VHGRYGHDSPIGSTVFSCDGYSRLSHEBPLCEKMMWSHPLPTCDALCGGVDRPSG 1039
693 TILSPGPDYPPNNLCTWIIETSHGKGVFTFTPHLESCHDVLTLTENGSPFOPLRQL 752
1040 TILSPGPEYPPNLTCTWVDVTHGKGVFNFTPHLEBHDHDLTLTENGSPFOPLRQL 1099
753 TGSRLPPIISAGYNGTAAOVRFISDPSMEGNIITFSEXDYCEPEVPAVSIRKGL 812
1100 TGSRLPPIINAGYNGTAAOVRFISDPSMEGNIITFSEYNTLEPCDDPGIPQYGSIRGF 1159
813 QFVGVDLTTFSCPEYRLBGTARITCLGGRRLMSPLPCVARECGSVTGTQTLSPN 872
1160 NFGIGDPLTFSCSGYRLBGTSEITCLGGRRLVMSAPLPCVARECGSATNNEGILLSPN 1219
873 FPNVYNNNHGCIYSIQTPGKGIOLKARABELSEGDVLYKYYDGNNSARLLGVSHSMM 932
1220 YPLVYNNNHGCIYSIQVQAGKGINISARTFHLAQGDVLKTYDGDKTTHLLGATGASMR 1279
933 GVTLNNTSSSLMTDFITDAINTSGFELHSFELICEBPGPKFYKXVHDESGFSS 992
1280 GLTISSTSNQMLMEFNSDTRGIDGFLVTSFELHSCEBPGCLIPQFYKISDQHFAGST 1339
993 VSPSCDPYSLRGSEBELCLSGERRTRDRLPTCVARECGTVRGEVSGOVLSPGYAPYE 1052
1340 IYGCNPGYTLHGSILKMTGERRANDYLPSCIAACGGRFKGESGRLLSPYPPPYD 1399
1053 HNLNCTITAEAGCTIGLHPLVFDTEBVADVLRIMGPVBSGVLLKELSGPALPKULHS 1112
1400 NNLRCMMIMIEVDENIYSIQFLAEDTASHDIIRVMGPPENMLLKEISGSLIPRGIHS 1459
1113 TPNVSVLQFSDPFTSQGFALIOFSVSTARSNDPGIPKONGSGDSWEAGDSVTFPCDP 1172
1460 TNLVITVITQFDITFYSISGFAIQFSSVATACADPGVPMNGITRAGDEREGDITVFCDDP 1519
1173 GYALQGSABEISCVKIEENRFPMQSPPTCIAPCGGDLTGPSGVLLSPNYPAPYPGKXCDW 1232
1520 GYELQGSERITTCIQVENRYFMQSPSPVICAPCGGNLTGSSGFLSPNYPHYPSRBCDW 1579

```


Qy		2313	VSSISVTHSGMRRLIFETFOYQFOAOLMLCDPEYYTGTGQRVLRCQANGKMSLGDSTPTCRI	2372
Dd		2660	INSFLEIHERMRLVNGSHYEYTKTFVSCDPPCHGLGPASIECLPNGTSMWNERPYOI	2715
Qy		2373	ISCGELPIPDNHRIGTLTSVYGATAIFCSNSGYTLVGSRVECMANGLMSGSEVCLAGH	2433
Dd		2720	ISCGELPTEPNANKIKIGTQTSYGSTAIFTCIDGLFMLVGSAVHECCSLSGLSMBSETRKLAGH	2779
Qy		2433	CGPREPIIVNHGINGENYSTRGSVVYQCNAHPILLIMSVRIQQODHMWSGKTPEPCYPITCG	2492
Dd		2780	CGIPELLIVAGQVIENGVGRDVIYVCQNPFRLIGSSVRICQQODHMWSGOPLSCVPVYSCG	2839
Qy		2493	HGNPNVNLTOQNOFLNDVVFVFCNPGYMAAGARSOCCLASGGOWSDMLPCRILINCNDP	2552
Dd		2840	HGSPFYLGRTSNGNFNDVVTFSNCIGLYMGFPFKACQOANRQKSNHPYPCCKVYNCSDF	2899
Qy		2553	GHOEHSVRO--VHASGPHRSFGTTYSRCNNHGFFLLGTPLYSCQGDTWRDRPRPOCLL	2609
Dd		2900	GIPIANSKRSKSLIENH---FTYGVAVFYDDCNPGFYELFGSSVLLICPQMGDKPLPECIM	2955
Qy		2610	VASCGHPGSPHHOMSGDSYTVGAVVRYSCTGKRTLNGSTMCLDGHWTSLSLHCGSTS	2669
Dd		2956	IDCGHPGVPPNAVLSGEKYTFGSIYVHNSCTKRSLLGSSSRTCOLNGHWSGSQPHCSGDA	3015
Qy		2670	VGVCGDPGIIPAHGIRLGDSPDGTVMRFSCBAGHYLNRSSERTCOANGSWSGSQPECGVI	2729
Dd		3016	TOTCGDPPGTPGHSGHSESNFRKTSIVRYACDTGYILHGSBERTCLANGSWTGROPECKAV	3075
Qy		2730	SCGNNGTGSNMARVVSDDLVSSTIYECRGYATATGLSHRCVNGTWGSDPECLVIN	2789
Dd		3076	QCGNNGITFANGNGVFRIDGTYTSSSVIYSCMEGYLLSGPSVRAQCPANOTWSGTLFNCCIIS	3135
Qy		2790	CGDPGIPANGRLGDNDFRYNKTVLYOCVPGYMESHSHVSYLSTCYKDRTMNGTKRVCKALM	2849
Dd		3136	CGDPGIPANGRLRYGDDVYVGGNVSYMOPGYTWELNRSRIARTCTINTWGSMTPCAAYT	3195
Qy		2850	CKRPPELLINGKVSGSDFMWGSSVITYACLBEYQLSLPVLFTCEBGNGSWGTELPOCFPVFCG	2909
Dd		3196	CPTPQJISNGLEBTINPDMPGISIYSIPGEILSFPAVLTCVGNGTWSGEPVQCLPKFCG	3255
Qy		2910	DGVPSRRGRREDRFGFSYRSYSFSCHPPLVLYGSPRFGCSDGWSGTOSCIDPULTTC	2969
Dd		3256	DPGIPAGCKRKRGKSFIYQSSEVSBFCNPFLVGSSTRICADGTWSSGSHPICLEPIQTSC	3315
Qy		2970	ADPGVPOFGIONNSQGYOVGSYVLFRCKQKYLLOGSTTRTCLPNTLTWSGTPDCVPHCR	3029
Dd		3316	ENPGVPRHGSQNNTFGFQVGSVAVOFHCKKHLLQGSTTRTCLPDLTWSGIQPEICIPHSC	3375
Qy		3030	QPETPHANYGALDLPBMGYTLI 3052	
Dd		3376	QPETPALNAVGMIDLPSHGTYLI 3398	

RESULT 5
CSM3_MOUSE

ID	_CSM3_MOUSE	STANDARD;	PRT; 2796 AA.
AC	080779; OBBVU0; Q9D588;		
DT	29-MAR-2004 (Rel. 43, Created)		
DT	05-MAR-2004 (Rel. 43, last sequence update)		
DT	05-JUL-2004 (Rel. 44, last annotation update)		
DE	CUB and eushi multiple domains protein 3 (Fragment).		
GN	Name=Csm3; Synonyms=Klaa189;		
OS	Mus musculus (Mouse)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBT_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	TISSUE=Brain;		
RX	MEDLINE=22579291; PubMed=12693553;		
RA	Okezaiki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,		
TA	Nakajima D., Nagase T., Ohara O., Koga H.;		
TI	"Prediction of the coding sequences of mouse homologues of KIAA gene:		

RT II. The complete nucleotide sequences of 400 mouse KIR4-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.
RL DNA Rep. 10:35-48(2003).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND SEQUENCE OF 2379-2796 FROM N.A.
RP (ISOFORM 1).
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata, and Testis;
RX MIML=C57BL683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kanakawa T., Adachi J., Bono H., Kondo S.,
RA Nishida I., Osato N., Saito R., Suzuki H., Yamanaoka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schmitt L.M., Knapin A., Matsuda H., Baralov S., Beisel K.W.,
RA Blake J.A., Brad D., Brucic V., Chochia C., Corbani L.E., Cousins S.,
RA Dalla E., Digrant T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Gough J.,
RA Grummond S., Guscinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Karni A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lennard B., Lyons P.A.,
RA Maglocz D.R., Matsals L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nmatu K., Okido T., Pavan W.J., Petrea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sanderlin A., Schneider C., Sempke C.A., Setou M., Shinada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Vercato R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmfing L.G., Wyshew-Borle A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinici P., Hayatsu N.,
RA Hironaka-Tsukikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Atakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yamauchi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RC -1- SOURCE/LOCUS LOCATION: Type I membrane protein (Potential).
CC -1- SOURCE/LOCUS PRODUCTS:
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=080779-1; Sequence=Displayed;
CC Name=2;
CC IsoId=080779-2; Sequence=VSP_009053, VSP_009054, VSP_009055,
CC VSP_009056;
CC Note=No experimental confirmation available;
CC -1- SIMILARITY: Belongs to the CSMD family;
CC -1- SIMILARITY: Contains at least 10 CUB domains;
CC -1- SIMILARITY: Contains at least 25 Sushi (CCP/SCR) domains.
CC -1- CAUTION: Ref.2 (BAC37116) sequence differs from that shown due to
CC multiple frameshifts from position 2763.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK125672; BAC65849.1; -
CC EMBL; AK015672; BAB2924.1; -
CC EMBL; AK078076; BAC37116.1; ALT_FRAME.
CC MGD; MGI:1922193; 4930500N14R1K.
CC InterPro; IPR000859; CUB.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00431; CUB; 10.
CC Pfam; PF00084; Sushi; 25.
CC PROSITE; PS01180; CUB; 10.
CC PROSITE; PS50923; SUSHI; 25.
CC Alternative splicing; Repeat; Sushi; Transmembrane.
CC NON_TER 1

FT	DOMAIN	1	52	Sushi 1.
FT	DOMAIN	54	162	Sushi 1.
FT	DOMAIN	167	224	Sushi 2.
FT	DOMAIN	226	336	Sushi 2.
FT	DOMAIN	339	398	Sushi 3.
FT	DOMAIN	400	508	Sushi 3.
FT	DOMAIN	511	571	Sushi 4.
FT	DOMAIN	573	682	Sushi 4.
FT	DOMAIN	685	745	Sushi 5.
FT	DOMAIN	747	855	Sushi 5.
FT	DOMAIN	858	919	Sushi 6.
FT	DOMAIN	921	1034	Sushi 6.
FT	DOMAIN	965	1026	Sushi 7.
FT	DOMAIN	1028	1136	Sushi 7.
FT	DOMAIN	1139	1198	Sushi 8.
FT	DOMAIN	1200	1308	Sushi 8.
FT	DOMAIN	1311	1370	Sushi 9.
FT	DOMAIN	1372	1483	Sushi 9.
FT	DOMAIN	1482	1543	Sushi 10.
FT	DOMAIN	1545	1656	Sushi 10.
FT	DOMAIN	1656	1778	Sushi 11.
FT	DOMAIN	1719	1780	Sushi 12.
FT	DOMAIN	1781	1845	Sushi 13.
FT	DOMAIN	1846	1903	Sushi 14.
FT	DOMAIN	1904	1961	Sushi 15.
FT	DOMAIN	1962	2019	Sushi 16.
FT	DOMAIN	2020	2081	Sushi 17.
FT	DOMAIN	2082	2139	Sushi 18.
FT	DOMAIN	2143	2200	Sushi 19.
FT	DOMAIN	2201	2259	Sushi 20.
FT	DOMAIN	2260	2319	Sushi 21.
FT	DOMAIN	2320	2377	Sushi 22.
FT	DOMAIN	2378	2435	Sushi 23.
FT	DOMAIN	2439	2497	Sushi 24.
FT	DOMAIN	2498	2557	Sushi 25.
FT	CARBOHYD	125	125	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	251	251	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	285	285	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	330	330	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	439	439	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	695	695	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	750	750	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	868	868	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	940	940	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1018	1018	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1108	1108	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1244	1244	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1375	1375	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1380	1380	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1413	1413	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1584	1584	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1626	1626	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1800	1800	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1831	1831	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	1951	1951	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2021	2021	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2041	2041	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2188	2188	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2247	2247	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2256	2256	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2283	2283	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2297	2297	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2307	2307	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2365	2365	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2453	2453	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2611	2611	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2618	2618	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2701	2701	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2707	2707	N-linked (GlcNAc . .) (Potential)
FT	CARBOHYD	2716	2716	N-linked (GlcNAc . .) (Potential)
FT	VARSPIC	1	220	Missing (in isoform 2).
FT	VARSPIC	221	223	FTCD -> MAK (in isoform 2).

FT FTID=VSP 009054.
 FT AECGASATNNNEG -> GTMSAFLCLYV (in isoform
 FT 2).
 FT /FTID=VSP 009055.
 FT Missing (in isoform 2).
 FT VARSPLIC 410 2796 /FTID=VSP 009056.
 FT SEQUENCE 2796 AA; 305500 MW; D62B2F1E758F79DD CRC64;
 Query Match 53.0%; Score 8997.5; DB 1; Length 2796;
 Best Local Similarity 59.0%; Pred. No. 0;
 Matches 1534; Conservative 423; Mismatches 564; Indels 77; Gaps 3;
 QY 488 NGKRGDLDLGSSISFLCDGEGFLTGQSGEITTCVLEKSGVVMNSAVLRCAPCGHLTS 517
 Db 1 NARRGDNFOLGSSISIVICEEGFIKTQGETETITCILDGMKYMGGPIPRCAPCGHFA 60
 QY 518 PSGTILSPGMPGVEYDALSCAMVLEAOGYRIKITTDFRKTENVYDTLEVRDGTYSAPL 577
 Db 61 PSGVILSGMPGVEYDNLNCEWVLEAEGHSIKITFERFOTELNVYDVLEVHDGPNLLSPL 120
 QY 578 IGVYHGTQVPOFLISTSNVYLIFSTDKSHSDIGFOLRYETITLQSDHCLDPGVNRCOR 637
 Db 121 LGSVNGTQVPOFLRSSNFYLLFTTDSNSNNGFKIHESVYVNTYSCLDPLVHGR 180
 QY 638 HGNDPVCALVTEGCDGTYTSDGEPLECEPNFQWSBALPSCCALCGGFIQSSGTLSP 697
 Db 181 YGHDFISGISTVSFSCDQGYRLSHREPLCEKNHWSHPLPTCDALCGDVRGPGSGLTSP 240
 QY 698 GPPPYFNNLNCWTYIETSHGKGVFTFPHLESGLDYLLITENGSPFPLAQLTGSRL 757
 Db 241 GYFEPYNSLNCITVTVTHKGVQFNFPHLDHNDHYLLITENGSPFPLAQLTGSRL 300
 QY 758 PAPISAGLYGNFTAVQVFFISDPSMSYEGFNITSEYDLECEPEVPAVYSIRKGLQGVG 817
 Db 301 PSTINAGLYGNFTAVQVFFISDPSMSYEGFNITSEYDLECEPEVPAVYSIRKGLQGVG 360
 QY 818 DTLFESGPGVLEGTARITCLGGRRLNNSPLPRCAEAGSVTGTQGLTSPNPNVY 877
 Db 361 DTLFESGPGVLEGTARITCLGGRRLNNSPLPRCAEAGSVTGTQGLTSPNPNVY 420
 QY 878 NNNHECTYSIQTPGKGIQKARAFELSEGDVLYKVDGNNNSARLLGVFSHSMGVTLN 937
 Db 421 ENNHECTYSIQTPGKGIQKARAFELSEGDVLYKVDGNNNSARLLGVFSHSMGVTLN 480
 QY 938 STSSSLMLDFTTAENTSKGFEILHSSFEILKCEDPQPKGYKVADDEGHFAGSSVSFSC 997
 Db 481 STSNQMLWFENSDSEGDGFLVYTSFELSHCEDPQPKGYKVADDEGHFAGSSVSFSC 540
 QY 998 DPYSILRGSEBELCLSGERTTMDRPLPTCYABCGGVTVRGEVSGQVLSRGYAPYEHNLNC 1057
 Db 541 NPQYITLHSSSLKCMGTERRAMDYPLPSCIAECGGRFKBESSGRILSPGYPPYDNLNRC 600
 QY 1058 IWTIEBAGCTTGLAHLFVDETEVHDVLRIMDGPVSEGVLYLKELSGALPKDLHSTPNVY 1117
 Db 601 MMWIEVDPGNIVLQFLAFDTEASHDILRWMDGPPENEMLLKEVSSSLIPDGHSTINIV 660
 QY 1118 VLOFSDFTSKQGFALIOFSVATATSCNDPGLPONGSRSGDSWEAGDSTVPCQDPYALQ 1177
 Db 661 TLOFDDFYISKSGFALIOFSVATATSCNDPGLPONGSRSGDSWEAGDSTVPCQDPYALQ 720
 QY 1178 GSAELSCVXLENRFQWQSPPTCIACCGDLTGPGSVILSPNYPBPYPRGKCDMKVYVS 1237
 Db 721 GQGRICICQVENRYFQWQSPPTCIACCGDLTGPGSVILSPNYPBPYPRGKCDMKVYVS 780
 QY 1238 PDYVILAVNINFLLEGYDLFHIYDGRDSLPLIGSFYGSQULPGRHIESSNSLFLAFRSD 1297
 Db 781 TTYVILSLAFISFISIBENYDFLYIYDGPDSNPLIGSFYGSQULPGRHIESSNSLFLAFRSD 840
 QY 1298 ASYSNAGFYIDYENPRESCEPDSIKXNTRVGSULKAGSSVTTYCHAGYEVEGTSTLSC 1357
 Db 841 GSVSYTGFLLEYKAKLRESCEPDSIKXNTRVGSULKAGSSVTTYCHAGYEVEGTSTLSC 900
 QY 1358 ILGPGDKPVMNNRPVCTAPCCGQYVGSQVLSBNYPQNTSQQICLAYFTVPRDYVF 1417

Db 901 IMEDDPRPGWNRVLPSCAPCCSRSTGSEGTVLSBNYPKXVSDHNCVYSIAVPEKEF--- 957
 QY 1418 GQFAPFHTALNDVVEVHDGSHQSHRLSSLSGSHTESESLPLATSNQVLLKFAKGLAPAR 1477
 Db 958 -----AVPRTSITQSSVPEPRFGRIRIGNFAVGSILVLFECNPOXYILHGRALRCERTV 1010
 QY 1478 GFHFVTVQAVPRISATQSSVPEPRYGRKLGDSFVGAIYRPECNGYALQSGPEIECLPV 1537
 Db 958 -----AVPRTSITQSSVPEPRFGRIRIGNFAVGSILVLFECNPOXYILHGRALRCERTV 1010
 QY 1538 PGALQOMNSAPTCVAPCCGNTLTERGTLISGFEPRYNSLNCNWKIVYVPGAGIQOV 1597
 Db 1011 PMSLAOMNSLPTCIYVPCGGLITKRGTLISGVEPEPDNNLNCNWKITVPGAGIQOV 1070
 QY 1598 VSFVTEQWMSLEVDGADNTVYMLSGSGTVPPLNLNSTSQLYLHFYSDSVSAAGPH 1657
 Db 1071 VSFATEHMDSLDFPDGDDNNAPRLGYSVGTIIPLHLNSTSNLNLINQOSDVSAGPH 1130
 QY 1658 LEYKTVGLSSCEEPVAPNSGVKTERVYLVNDVVSQCEPGLAQHAHISCMPTVRBN 1717
 Db 1131 LEYTAIGLDSCEBPQTPSSGIKVGDRVMVGVSFQCDQGYSLQSHSHITCMPGVRRBN 1190
 QY 1718 YPPPLCIAQCGTVEEMGVILSPGPNYDPSNMDCKIALPVGFGAHIQPLNSTEPN 1777
 Db 1191 YPIPLCIAQCGAMDSFGSVILSPGPNYDPSNMDCKIALPVGFGAHIQPLNSTEPN 1250
 QY 1778 HDYIIRNGPYETSHMGRFSGSELPSSLSSTHETTVYFPHSDHSONRPGKLEVOAYEL 1837
 Db 1251 HDYLEVRSGSSEISVIGRLSGPQPSLSFTTHETSLYFPHSDHSONRPGKLEVOAYEL 1310
 QY 1838 QCEPPEPFANGIYRGAANYNGQSTPECLPGYQLTGHPVLTQGHTRAMNDHLPKCEV 1897
 Db 1311 QCEPPEPFANGIYRGAANYNGQSTPECLPGYQLTGHPVLTQGHTRAMNDHLPKCEV 1370
 QY 1898 PCGNITSNNGTVVSPGFPSPYSSSODCVMLITVPIGHGVLANLTLQTESGSPFITWD 1957
 Db 1371 LCGGNITMNGTITVSPGFPSPYSSSODCVMLITVPIGHGVLANLTLQTESGSPFITWD 1430
 QY 1958 GPQCTAPPLGVFTSMKAKTVYQSSNOVLLKFRHDAATGGIIFALFASAYPLTKCPPTIL 2017
 Db 1431 GPQDMSPGIQFQSNITLLESYVSTSNQIILKFHSDFTTSGFPVLYAHYQLRVCQPPPV 1490
 QY 2018 PNAEVTENEEFNIGDLYRRCPLGFTLVNGELLTCLGTYLQREBPPLCEVNCPTNEL 2077
 Db 1491 PNAELTREDDEFELGDIIRQCLPGFTLVNGELLTCLGTYLQREBPPLCEVNCPTNEL 1550
 QY 2078 LTTSTGYILSGSYPGSYPOFQTSMLVREPDYVNLSTVEYFLSEKQYDFEFLPDGSGQ 2137
 Db 1551 RLDETGYILSGSYPGSYPNLQMCAMSISVEKGNVISMFEFQTEKXFDVLQYVDGPNIQ 1610
 QY 2138 SPILKALISGNSAALITVSSNSNYLRLMSDDHAYNRKGFIRISAPYCSLPRAPLHGFIL 2197
 Db 1611 SPVLISLSGDSIAFENVTSNGHEVFLQMSADHGNNKGFIRIRIAYFCSYPRPSPHYII 1670
 QY 2198 GQTSSTOPGGSIHFGCNAGYRLVGHSMALCTRHOGYALWEALPLCOALSQGLPEAPXNG 2257
 Db 1671 SQTGQGLNSVVRNACDGRFLVGRSSAVCKKSSYGHSMDAPVACALSCGIPKATNG 1730
 QY 2258 MVEKKEYTGTKAVYSCSEGYHLQAGAATAECLDTGLMSNRNVPQCVVTCPDVSSIS 2317
 Db 1731 GILTTDYLVTGRVTVFENDGYRLSKELTATQSDGTGMSNNHKTRCVAVVTCPSINSFP 1790
 QY 2318 VEHGRWRLIFETQYQFOQAOLMLCDGYYVYTGQVRVRCQANGKMSLGDSTPTRIISCGE 2377
 Db 1791 LDBGRWRLIVGSHYKTKKXVGVSCDPGQYHGLGASIECLPNGTWSRTERPYQIISCGE 1850
 QY 2378 LPIEPNGHRIIGLTVYGATALFSCNAGYTLVGSRVCECMANGLMSSEYKCLAGHGCTPE 2437
 Db 1851 LPIPPNGNRIIGLTVYGATALFSCNAGYTLVGSRVCECMANGLMSSEYKCLAGHGCTPE 1910
 QY 2438 PLYNGHINENYYSRQSVYQCNAGFRLIGMSVRIQODHMSGKTPFCVPTTCGHPNP 2497

Db 1911 LIVNGQVIGENVYEDTVVYOCNPGFRLIGSSVRIQQODHNWSQLPSCVPEVSCGHPSP 1970
 2498 VNGLTGQNGFNLDVYKFCNPGVYBAGARSQCLASOWSMDLPTCIIINCTPRGHEN 2557
 Db 1971 IYGTSGNGFNNDVYVPSCNIGYLMGPTKACQCANQWHPVPCVKNVNCSDPGIPAN 2030
 2558 SVRQ---VHAGSGPHRFSGTIVSYRCNNGFYLLGTVPVLSGCGDGTDRPRQCLLVSCGH 2614
 Db 2031 SKRESKIEHGN-----FTYGTVFYDCNPGYFLPSSSVLICQPNQMDPELECIIMIDCGH 2086
 2615 PGSPPHSGMGSSTYVGAIVNRYSCIGKXTLVGNSTRMCGLDGHTGSLPHSCGTSVGVCG 2674
 Db 2087 PGIPPNVATSGEKTFCGCTVHYVSCGKRSLLGQASRTCOLNGHWSGSPHSCGDTTGTGCG 2146
 2675 DPGIPAHGIRLGDSDPCTWRFSCGACHVLRGSSERTCOANGSSGQPEPCGVISGPNP 2734
 Db 2147 DPGIPGHSGRQSDPRKYSIVRFACDITGILYGSSERTCLNSGWTGKQPECKAVQCGNP 2206
 2735 GTPSNARVVFSDGLVFSSSIYECREGYATGLSRHCSVNGTWTGSDPECLVINCGDPG 2794
 Db 2207 GTTANGKVFRIIDGTFSSSVIYSCLEGYILSGPSVROCTANGTWSGSLPNCITIIISGDPG 2266
 2795 IPANGRLRGDNFRYKNTVYTCVPGYMHSHRVSVLCTKDRYVNGTKPVCKALMCKPP 2854
 Db 2267 IPANGRLRGDNFRYKNTVYTCVPGYMHSHRVSVLCTKDRYVNGTKPVCKALMCKPP 2326
 2855 LIPNGKVVGSDFPMWSSVTYACLBGYQLSLPAVFTCEBNGSWTGEILPCFPVFCGDPV 2914
 Db 2327 QISNGRLEGTFNPDGFSISYICSAAGELSPFVAVLTCVNGTWSGVPCLCKFGDPDIP 2386
 2915 SRGRREDGFSYRSSVSFSCHPPLVYSGPRFCOSDGTSGTOPSCIDPTLTTCADPGV 2974
 Db 2387 SGGRREGSFTYQSEVSFSCSPFLVSSSTRLCOTDGTWSSGSHCIEPRTISCENGV 2446
 2975 POFGIQNNSGYQVGSYVLPFCQKGYLLQSGTTRTCLPNTLWSGTPDPCVPHKQPEPT 3034
 Db 2447 PRHGSQNNTFGFGVGSVVQFPHCKGHLQSGTTRTCLPDLTWSGIIQPECIPIHSCQPEPSP 2506
 3035 THAVGALDLPMSGTYLL 3052
 Db 2507 AHAHVGMGLDPSHGTYLL 2524
 RESULT 6
 072241
 ID 072241 PRELIMINARY: PRT: 1466 AA.
 AC 072241;
 DT 01-JUN-2003 (TREMBLREL. 24, Created)
 DT 01-JUN-2003 (TREMBLREL. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBLREL. 26, Last annotation update)
 DE SI: b21C3.1 (Novel protein similar to CUB and Bush1 multiple domains
 DE protein (CSMD)) (Fragment).
 GN Name=SI:b21C3.1;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Babbage A.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL732499; CAD60796.1; -.
 DR HSBP; 09UCV4; INZ1.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR001000; Glyco_hydro_10.
 DR InterPro; IPR000436; Bush1_scr_CCP.
 DR Pfam; PF00431; CUB; 9.
 DR Pfam; PF00084; Bush1; 8.
 DR SMART; SM00032; CCP; 8.
 DR SMART; SM00042; CUB; 9.
 DR PROSITE; PS01180; CUB; 9.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; UNKNOWN_1.

DR PROSITE; PS00923; SUSH1; 8.
 FT NON_TER 1
 FT 1466 1466
 SO SEQUENCE 1466 AA; 160456 MW; 45F45ED456FF472E CRC64;
 Query Match 36.4%; Score 6180.5; DB 2; Length 1466;
 Best Local Similarity 74.1%; Pred. No. 0;
 Matches 111; Conservative 170; Mismatches 181; Indels 37; Gaps 4;
 135 MDAHLRPGSITSPNFIQYDNNACWITLANSKYIKALFEEPDLERGYDTLVG 194
 Db 2 MCGQLGPGSGIITSPNFIQYDNNACWITLANSKYIKALFEEPDLERGYDTLVG 61
 195 DGGDDGOKTYLVYNSQACSDSPHTPSRIPE---SMGDIWRKMTVLEICRISSDA 251
 Db 62 DGSVIGDQRTVFEVH-----LSGTTTDLVYSTSHQWMLN----- 95
 252 RSGSVRSKPTSNVAVLADGTEIEQSCGDPGIPAYGRREGSRFHHDTLKFCOPAFE 311
 Db 96 -----FTDDTSSLGFRKVSVEEIDQGCCDPGIPAYGKRGITFRHGDLYFECLEPAFE 150
 312 LVGQXAITCOKNQWSAKKQCVFSCFPNFTSPSGVLSPNYPEDYGNHLHCWLLIARP 371
 Db 151 LVGKKKITCOKNQWSAKKQCVFSCFPNFTSPSGVLSPNYPEDYGNHHCWLLIARP 210
 372 ESRHILAFNDIVDEPODFVIYKDGATAEAPVLTGTFSGNOLPSSITSSGHVARIETPDH 431
 Db 211 ESRINLAFNDLSMEKQDFLSIKDQKAESEPTLSTFGSDVLPSTITSGHVARLEPLTDH 270
 432 STGRKFNITFTTFRHNECEDPDPVAVNGKFFGDSILQSSISFLCDGFLQTQSEETTC 491
 Db 271 TYDGRGNITFTTFRHNECEDPDPVAVNGKFFGDSILQSSISFLCDGFLQTQSEETTC 330
 492 VLKGSVAVNSAVLRCEAPCGGHLTSPSGTILSPGMPFYKDALSCAMVIAOQYPIKI 551
 Db 331 ILKGNVAVNNAVLRCEAPCGGHLTSPSGTILSPGMPFYKDALSCAMVIAOQYPIKI 390
 552 TFDKFTENVYDTELVADGRTYSAPLIGYHGVQVOPQLSTSNVYLLSTDSKSHDID 611
 Db 391 IFDKFTENVYDTELVADGRTYSAPLIGYHGVQVOPQLSTSNVYLLSTDSKSHDID 450
 612 FOLREYITLQSDHCHDPGIPVNGQRHGNFVYVALVTFSCDSGYTLSDSEPLECEBNFQ 671
 Db 451 FRIREFILOQSDHCHDPGIPVNGQRHGNFVYVALVTFSCDSGYTLSDSEPLECEBNFQ 510
 672 WSRALPSCDALCGFGIQSSGTTILSPGPDPPYNNLNCWIIETSHKGVYFTFTFHL 731
 Db 511 WSRALPSCDALCGFGIQSSGTTILSPGPDPPYNNLNCWIIETSHKGVYFTFTFHL 570
 732 SGHLYLITENGSTQRLQRLTGSRLPAPISAGLYGFTQVRIISPSMSYBGFNTTFS 791
 Db 571 SPDHLLVTENGSSQPLMLTGSTLPPISAGLYGFTQVRIISPSMSYBGFNTTFS 630
 792 EYDLEPECEBEVPAYSIRKGLQFGVGDTLTFSCPPGRYLRGTARITLGGRRRLMSPLP 851
 Db 631 EYDLEPECEBEVPAYSIRKGLQFGVGDTLTFSCPPGRYLRGTARITLGGRRRLMSPLP 690
 852 RCVAECNSVYTGQTLISPNFPVYNNNHCIYSIQTOPQKGIQLKARAFELSEGDVLE 911
 Db 691 RCVAECNSVYTGQTLISPNFPVYNNNHCIYSIQTOPQKGIQLKARAFELSEGDVLE 750
 912 VYDGNNSAALLGVFSHSEMAGTILNSTSSLMIDFTDADNNSKGEELHFPSSFEILICE 971
 Db 751 VYDGNNSAALLGVFSHSEMAGTILNSTSSLMIDFTDADNNSKGEELHFPSSFEILICE 810
 972 DPGPKFGYKVDHGEHAGSAGSVSPSCDPSGLKSGSEELLCLSGERRRTWDRPLPCVACG 1031
 Db 811 DPGPKFGYKVDHGEHAGSAGSVSPSCDPSGLKSGSEELLCLSGERRRTWDRPLPCVACG 870
 1032 GYVAGSVSGVLSFGYAPYEHNLNCIWTIEAGAGCTIGLHFLVFTDEBVHDLRIWDGP 1091
 Db 871 GKIDESMGRLISPGYAPYEHNLNCIWTIEAGAGCTIGLHFLVFTDEBVHDLRIWDGP 930

```

QY 1092 VESGVLLKELSGPALPKDLHSTFNSVLIQSTDFSTKQGFALQPSVSTATSCNDPGLI 1151
D 931 QGGVTLRELSTGSLPPDLHSTFNSVLIQSTDFSTKQGFALQPS---ATSCNDPGLI 987
QY 1152 NSRSGDSWEADSTVFCQDDPGYALQGSALISCVKIENRPFWQSPPTCIAPCGGDLTGP 1211
D 988 NSTRIGDSREPDHVLFCQDDPGYALQGSALISCVKIENRPFWQSPPTCIAPCGGDLTGP 1047
QY 1212 SGVILSPNYPPEPPEKCEKDMKVTSPPVIALVINFLNPGVDFLHYGRLSLPLI 1271
D 1048 SGLILSPPEPPEPPEKCEKDMKVTSPPVIALVINFLNPGVDFLHYGRLSLPLI 1107
QY 1272 GSFYGSQPLGRTESSNSLFLAFRSDASVSNAGFVIDYENRPSCEPDGSIKNGTVGS 1331
D 1108 GSFYGTDPDRTESSNSLFLAFRSDASVSNAGFVIDYENRPSCEPDGSIKNGTVGS 1167
QY 1332 DLKLGSSVTVYCHGVEYEGTSTLCILPGPKPVMNPRVCTAPCGQYVSGDGYVLS 1391
D 1168 ELKLGATVTVYCHGVEYEGTSTLCILPGPKPVMNPRVCTAPCGQYVSGDGYVLS 1227
QY 1392 PNYPONTYSGQICLYFTVTPKDYVVFQFAFHTALNDVVEVHDGSHSLSLSGSH 1451
D 1228 PGYPNYSNGRTCLYSVVPDYVVFQFAFHTALNDVVEVHDGSHSLSLSGSH 1287
QY 1452 TGESLPLATSNQVLIKPSAKGLAPRGHFFVYQAVPRTSATQSSVPEPRYKRLGSDFS 1511
D 1288 TGESLPLATSNQVLIKPSAKGLAPRGHFFVYQAVPRTSATQSSVPEPRYKRLGSDFS 1347
QY 1512 VGAIVRECEKNGVALQSGPEIECPVPGALAQMWVSAPTCVPGGMLTERRGTLISPGF 1571
D 1348 VGAIVRECEKNGVALQSGPEIECPVPGALAQMWVSAPTCVPGGMLTERRGTLISPGF 1407
QY 1572 PEPILNSLNCVWKIVVEGAGIQIQVVSPTQNMDSLEVFADGANTVTMLGFSFGTIV 1630
D 1408 PEPILNSLNCVWKIVVEGAGIQIQVVSPTQNMDSLEVFADGANTVTMLGFSFGTIV 1466

```

RESULT 7

```

Q9ES77 PRELIMINARY; PRT; 3567 AA.
AC Q9ES77;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-MAR-2004 (Tremblrel, 26, Last annotation update)
DE Polymorphism precursor.
GN Name=Polymorphism.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH/Hensic;
RX MEDLINE=20517255; PubMed=11062057;
RA Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA Romeo P.-H., Vigon I.;
RT "Polymorphism: a secreted protein with pentraxin complement control
RT protein, epidermal growth factor and von Willebrand factor A
RT domain."
RL Biochem. J. 352:49-59(2000).
DR EMBL; AF206329; AAG32160.1; -.
DR HSSP; Q9UCV4; INZI.
DR MGD; MGI:1928849; Polymorphism.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR009030; Growth_factor.
DR InterPro; IPR003410; Hyaluron.
DR InterPro; IPR001759; Pentaxin.

```

```

DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF_7.
DR Pfam; PF07645; EGF_CA_2.
DR Pfam; PF02494; HVR_2.
DR Pfam; PF00354; Pentaxin; 1.
DR Pfam; PF00084; Sushi; 33.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PR00010; EGFBLDOD.
DR PRINTS; PR00895; PENTAXIN.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD0002153; Pentaxin; 1.
DR SMART; SM00032; CCP; 34.
DR SMART; SM00179; EGF_CA; 6.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 6.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS00026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 6.
DR PROSITE; PS00825; HVR; 2.
DR PROSITE; PS00923; SUSHI; 34.
DR PROSITE; PS00234; VWFPA; 1.
DR EGF-like domain; Signal.
FT SIGNAL 1 17
SQ SEQUENCE 3567 AA; 387393 MW; 8FBA8276E12293E5 CRC64;

```

```

Query Match 11.0%; Score 1864.5; DB 2; Length 3567;
Best Local Similarity 21.4%; Pred. No. 3e-97;
Matches 793; Conservative 376; Mismatches 1173; Indels 1359; Gaps 148;

```

```

QY 99 TCNEGYDQSGKRTICMKVSDMPFAMSDHRCVCAKMDALRLRPSGILTPNPIQYDN 158
D 466 TCNEGYDQSGKRTICMKVSDMPFAMSDHRCVCAKMDALRLRPSGILTPNPIQYDN 512
QY 159 NAHCWMIITLNPSSKVITKLAFFEEFDELRGYDTLTVGCGQDGDQKTVLYMQNACSDSPH 218
D 513 NAHCWMIITLNPSSKVITKLAFFEEFDELRGYDTLTVGCGQDGDQKTVLYMQNACSDSPH 520
QY 219 TPGSRIPSMGSDIMROKWTVLEICRDISSDASGVSYSKTSNAVELVAPGTEIEQG 278
D 521 TPGSRIPSMGSDIMROKWTVLEICRDISSDASGVSYSKTSNAVELVAPGTEIEQG 567
QY 279 SCGPFGIPAYARRRGS-----RHHGDTLKFCOPAFE-----LVGQKAITGCKN 323
D 568 SCGPFGIPAYARRRGS-----RHHGDTLKFCOPAFE-----LVGQKAITGCKN 627
QY 324 NQMAKKEGCVFS-----CFNFTSPSGV-VLSPNYEDYGNHLHCWMLLA 369
D 628 NQMAKKEGCVFS-----CFNFTSPSGV-VLSPNYEDYGNHLHCWMLLA 671
QY 370 RPESRHIAFDIVDERPQDFVLYKDGATAEAPVLCGFS-GNQLPSSITSSGHVAREFQ 428
D 672 RPESRHIAFDIVDERPQDFVLYKDGATAEAPVLCGFS-GNQLPSSITSSGHVAREFQ 706
QY 429 TDHSTGRGNNITFTFRHNECDPVPVNGKRFSGSLQSGSISFLCDGFLTGSS-E 487
D 707 TDHSTGRGNNITFTFRHNECDPVPVNGKRFSGSLQSGSISFLCDGFLTGSS-E 765
QY 488 TITCVLKEG-----SVYMNASAVLRCEAPCG-----GHUTSPSGTI 522
D 766 TITCVLKEG-----SVYMNASAVLRCEAPCG-----GHUTSPSGTI 825
QY 523 LSPGMPGFYKDA-----LSCAWVIEAQGPY-----IKITPRFRT 558
D 826 LSPGMPGFYKDA-----LSCAWVIEAQGPY-----IKITPRFRT 885
QY 559 EVNYDTLEVVDGR-----TVSAPLIGVHGT-----QVPOFLIST- 593
D 886 EVNYDTLEVVDGR-----TVSAPLIGVHGT-----QVPOFLIST- 945
QY 594 ---SNLYLYLFSRDKSHSDIGPOLAREYETITLQSDHC-----LDPGLPVNGQR----- 637

```

Dh 946 ETITRLKTLINKEMYS---FQLASETVVADNSLSLTERKAFLEFCRPGSVLRGMCVNC 1002
Qy 638 HGNDPFGVALVTFQSDGYTLSDGEPLCE-----PFWMSRALPBCCLCGFLOGS 690
Dh 1003 LGTSLSEHSTCECLMSGYDEBQLECKLCPPRTHTLHRSVSECKAQCKOGTYSS 1062
Qy 691 SG-----TI-----LSPG 698
Dh 1063 SGLTECSECPGLGYQPERGSRSCLLCPETTTTVKRAVDISACGVPCEVGFSSGLTFC 1122
Qy 699 FP---DFYPNN-----LNCTW-----IETSHG 718
Dh 1123 YPCRDYQPNAGKSFCLACPFYGTITTTGATSIITDCSSFSSTSAEESIIVPAPGHS 1182
Qy 719 KGVP-----FTPHTHL-----ESGHYLLITENGSTFOPLROLTG----- 754
Dh 1183 QNKKEVSSQVHECEFLNCHNSGTCCQAGRGVCLCPRG-----YTGKCESTDIDEC 1234
Qy 755 SRLPAPISAGL-----YGNFTAQVRF-----ISDFMSY--- 783
Dh 1235 SSLFC-LNGGICROVVGFTCECSLGYSQICBENINECISPLCLNKGTCTDGLASTRCT 1293
Qy 784 --EGFNITFSEYDLERCEPEVPAVSIKRGLOFGVGDTLTFSQ-----FPGYRLEGTA-- 834
Dh 1294 CVKGYMGVHCETDVNECQSSPLNNAVCKQVGG-----FSCCKCPGFLGTRCEKNVDE 1347
Qy 835 -----RITCLGGRRLNASSPLRCVABEGNSVTGTGTLNBNFVNPN----- 878
Dh 1348 CLSOPCONGATCKKQAGNSF-----RC--QCPAGFTGHCEP-----NINECOSNPCR 1392
Qy 879 NNHBCI-----YSIOTOPKGIQLKARAFELSEGDVLKVYDGNNSAALLGVFHSHEMW 933
Dh 1393 NQATCVDELNYSCKCQCPG---FSGHRCETEP---SGFNIDEVSSIIYGVVLIDG 1442
Qy 934 V-TLNSTSSLMW---DFIT-----DAENTS---KGFELHFSFELI--- 968
Dh 1443 VLPHTLHATTCAFMMKMSDVINYGTPISYALBDDKMDTSLTIDYGMWLVYVNGKEKITNCP 1502
Qy 969 -----KCEBPGT----- 975
Dh 1503 SVNDGIWHHIAITWTSTGAMRVYINGELSDGTGLSIGKAIPEGALVLQODKKEG 1562
Qy 976 -----PKF-----GYKVD 984
Dh 1563 FNPABSFVGSISQNLMDYVLSPOOVKLLASSCEBELSRGNVAMPDFLSGITGVKYDS 1622
Qy 985 EGHFA-----GSSVSPGCDPGSVLRGSEBELCLSGERRTMDP 1022
Dh 1623 SSMFCSDPSEGSVPHLRPASGNRKPKSKVSLPCDFGQVGNPVQCLN--QGQWTP 1680
Qy 1023 LPTCV-AECC--GTVRGEVSGOVLSPGYRPAHYENHNLNCTIEMACCTIGLHFL--- 1074
Dh 1681 LPHCERIRCGLRPALLENFGVSAEDFHAAGTIVTQ-----CTSGYVILGDSR 1726
Qy 1075 VFDEEYVADVLRINDGPVSEGVLLKELS----- 1102
Dh 1727 MFCIDN-----GSMNGISPSCLVDDECAVSDCEHASCINTNGSYVCSNPPYTGDKN 1781
Qy 1103 -----GPALPKDLSTFNSVVLQFSTDEFTSKOGFAIOFS-----VSTAT-- 1142
Dh 1782 CABPVKCAKAPENPENGHS-----SGEITYT--VGTAVTFPSDEGHELVGISTITCLE 1830
Qy 1143 -----SCNDPGIIPONGSRGSDWEAGDSIVFOCDPGYALQSAEISCVKIE 1188
Dh 1831 TGEWDRLRPSCEALISCGVPVPVENGVDGSAFTYGSKVYVRCDDGYTLISGDEESACIASG 1890
Qy 1189 NRPFWQSPPTC-IAPCGGDLTGPSGVILSPNVEPYPGKECDMKVTVSPDYIA---- 1243
Dh 1891 S---WSHSPVCELVKC-----SQBEDINNGK-----YILSGITY 1922
Qy 1244 LVFNIENLEPGYDLHAYDGRDLSPLIGSFYGSQLPGRISSSNSLSFLAERSDASVNA 1303
Dh 1923 LSISYSCENGYSL-----QGFELLECTASGSV-----DRAPSC 1957

Qy 1304 GFVIDYTENDRESCFDPGSIKNGTRVGSDLKLASSVYYCHGYEVEGSTLSCLIPDG 1363
Dh 1958 QLV-----SCGEPPIYKDAVITGTSNTTFGTAVYTCKEGTLLGPDTTIVC--QANG 2006
Qy 1364 KPVNNBRPVCTAPCGQVYGSDDVLSPPNYPONTYSGOICLYFVTVPKD--YVFGQFAF 1422
Dh 2007 K--WNSSNHQCLA-----VSCBEPVVDHAS-----PETHRLRFQDTAF 2043
Qy 1423 FHTLNDVBEVHDGSHSRLLSLSGSHTGESLPLATSNQVLKESAKGLAPRGHVF 1482
Dh 2044 YYCA-----DQYS-----LADNSQLICNAQGNVPPAG----- 2071
Qy 1483 YQAVPRTSAPQSSVPEPRYG--KRLGSDPSVGAIVRFGNSGVALQGSBEICLPVP 1539
Dh 2072 -QAVPRCIAHCEKRPVSVSILESVAKAPAGSVSPFCMBGFLVNTSKIECLR-- 2127
Qy 1540 ALAQMNVSAPTCVPCGANTLERGTILSPGFPYPLNSINCWKIIVPBGAGIQIOWS 1599
Dh 2128 -----GEMS-----BSPL--SVQC----- 2140
Qy 1600 FVTEQNMDSLEVPDADNTVTMLGSFSGTVPALLNSTSNQVLYHFSYDISVSAAGHLE 1659
Dh 2141 -----IPV----- 2143
Qy 1660 YKTVGLSSCEPRAVPSNGVGTGERLYVNDVVSFQCEBRYALQGAHISCMPTVRBNYP 1719
Dh 2144 -----RCBPPSIANGYPSGTINSGAVAAVASHKGFYIKKSKSTC--BATQWSKP 2194
Qy 1720 PPLICIAOCGGTVEBEGVILSPGPGNYPNMDCSWKIALPVGGAHIOPLNFSTERNHD 1779
Dh 2195 TPTC-----HP--VSCNEBPXYENGLBHTTGTTFSEARFQ 2229
Qy 1780 YIEIRNGPYETSRMNGRFGSSELPSLSLTSHTTTFHSDHSONRPFLEVOAYELO 1839
Dh 2230 C---NPGYXA-----AGS--FVFVCOANR---HMHSDAPLSTPL-----N 2262
Qy 1840 CPDPEBPANGVAGAGVNGVQSVTFECLPQVLTGHVULCOHNTNNMD--HPLPKCEVP 1898
Dh 2263 CGKRPPIQNGPLKESEFVGSKVOFCVNEGELVGDNSWTCQ--KSGKMSKRPK--VP 2319
Qy 1899 --CGGNTSSNGTVYSGFPSPYSSQDCWMLTVPIGHGVRLNLSLQTEPSGDFYTIW 1956
Dh 2320 TKCAEPPLLENQVLKE-----LASEGVMTTISCCKGHALQ--GSPVLKCLPSQ-----W 2368
Qy 1957 DGPOQTAIRLGVTFRMAKTVQSSNQVLLKTHRAATGATGIFAIAFSAYPLTK--CP 2013
Dh 2369 NG-----SFPICGMVLCP 2382
Qy 2014 PTLIPNAEVTTENEERFIDIVRYRCLPGFTLVGNELITLCKLGYLOFEGPPTC--EVHC 2072
Dh 2383 PPLIPFG-VPASSGALHFGSTVKYLCVDGFFLRKSPITLQADS--TWSSPLPFCVVEBC 2439
Qy 2073 PTNELTDSGTVLISQSPGSPYPOFCQSMVLRVREPYNLSLVEYFLSEKQYDEFEIFD 2132
Dh 2440 PQREIILN--GIHVGOLALSTLLTYC-----KRFEL-----VGNATTLC 2479
Qy 2133 GPSGQ---SPLLKALSNGVSAPLITVSSNSVYLWSSDHAYNRKGFKIRIYAPYSLP 2188
Dh 2480 GENGMQIGKRMCKPIE-----CPEP 2500
Qy 2189 RAPLHGFILQOTSPQPGSIHFGCMAGYRLVGHSMALCTRHPOSYHLMSEAIPICOLSLC 2248
Dh 2501 KEILNG--QFSSVSFOYQOTITTYFCDRKFRLEGPKSLTCLBTGD--WMDPBCDAIHC 2555
Qy 2249 GLPEAPKNGVAFGEKYVTGTAVYSCEGYNLOAGABATACLDGTGLMNRNVPPOCVPV 2308
Dh 2556 SDPQRIENGVEBADYTGAMITYSCTPRGQVLGHANQT--CBESG--WSSSS--PTCVPI 2610
Qy 2309 TC---PDV---SSISVEHGMRLIFP-----T 2329
Dh 2611 DCGLRPHIDFGDCTKYVDGQHFDQEDDMMEVRYLAHQHLEATAKALENTKESPAHAS 2670

```

Qy 2330 QYQFOAQLMLICDPGYVYTGVRVIRCOANGKSLDSTPTCRILSCGELPI-PPNGHRIG 2388
Db 2671 HFLYGMVMSYSCPEYELLIGIPVLLIQDEGTWN--GTAPSCSLIC-DLPAPENGFIHF 2727
Qy 2389 TLISVGATATFSCNSGYTLVGSRVBCANGLMSGEVRCIAGHGTREPIVNGHNGEN 2448
Db 2728 TQITWGSAAQYSCKGHILIEGSHLRLCLONKQMSGTVPRCEALISCKRNPLMNSIKDD 2787
Qy 2449 YSRGSSVYVQCNAGFRLIGMSVRIQODHNSGKTPFCVPLITCGHPGVNGLQGNQFN 2508
Db 2788 YSLDELIVYEDCSGITLMSKKTQENRDMGDHEPMLCPVDCCSPVPTNGRVKGEYTT 2847
Qy 2509 LNDVVKFCVNCPEYMAAGARSQCLASGQMSDMLPTCRILNCTDGHQENSRYQVHASGPH 2568
Db 2848 FQKEITVSCREGEFIEGARSRLCLTNGSMGATPSCMVRCPAPQPVNGV---AAGLD 2903
Qy 2569 RPSFCTVSYCNHGFYLTGTIVLSCCDGTWDRPRPCLLVSCGHGSPHSGMSGDSY 2628
Db 2904 -YGFKEVAFHCLBSYVLAQAPRLTQSGNWTDAEVPVCKPATCGPPADLFQGFPPNGFSF 2962
Qy 2629 TVGAVVRYSCIGKRTLVGNSTRMCGLDGHWTSI.PHC-----SGTSVGV- 2672
Db 2963 YHGHGIQYQCFYGYKLNPNRRLCLPNMSWSSSPCLPCRCSTPTIIQGTINATIDJGCG 3022
Qy 2673 -----CGD-PGI PAHGIRLGDSPDGTV 2694
Db 3023 KTVQIECFKFKLGLSEITCDANGOWSDVPLCEHAQCGPLPTIP-NAIVLEGLSEDNV 3081
Qy 2695 MRFCSEAGHVLRGSSERTCOANGSMGSGQPECGVISCNGTPTNSA----- 2740
Db 3082 VTYGCRPYTWGSSDLCTEKAIMSQPYPCBFLSCGPPPTVANAATGEAHTYESKVK 3141
Qy 2741 -----RVVF-SDGLVFSSSIYV 2756
Db 3142 LRLCEGYVMSDITDTFTQODGHWVPERITCSPKCKVPSNMRI RHGDPQVNGQSV 3201
Qy 2757 ECRGYVATGLSHHCNVGNTWG--SDPRLVINCDBPGLPANGRLGNDPRYKNTVY 2814
Db 3202 SCABGFHEGVMNSTCPDPDGPWPFSDSCIPVYCHPSPSAPHGVSAGKHSFGSTIYV 3261
Qy 2815 QCVPGVMESHRYVLSCTKDRITWNGTKPYCKALMKCPRLIPNGKVGVSDFMGGSSVYV 2874
Db 3262 QCDPGYTLFEGNRRI--COENRQMSGGVANCRNRCCTPAPFNGRAVLENTTSGPSLIF 3319
Qy 2875 ACLEGYLSPAYFTTCGNSGWTGELFQCPVFCGDPGV-PSRRREDRGRFSYRSSVSFS 2933
Db 3320 SCHRGYTLBGSPEAHCTANGTWNHLTFLCKPNCPVFPVIBENAVLSEKEFYVDQVSLK 3379
Qy 2934 CHPLVLVGSPPRFCCGSDGTWSTGTPSCIDPILITCADPVPFGIQQNSOG--YVVGST 2991
Db 3380 CREGFLKNGVITCSDETHWTNARC--EKISCGPSHVENAL--ARGVYQYQDM 3433
Qy 2992 VLFRCQKGYLLQGSTRTTCLPNLTWSTGTPDCVHHCRQEPETHANVALDLPS----- 3046
Db 3434 ITYSYCSGYMLEGSLRSVCLENGTM--TPSPVCRAVCRP-----CQNGVCGQRPNACSP 3487
Qy 3047 ---MGYTLILPARRASPRVAPSTAPARRMAAGARPSAW 3084
Db 3488 DGMWGRIC-----EERPICPLNGRCVAPYQDCDCCPTGM 3522

```

```

RESULT 8
Q16744 PRELIMINARY; PRT; 2489 AA.
AC Q16744;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
OS Complement receptor 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94065175; PubMed=8245463;
RX VIK D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL and sequence of the coding region unique to the S allele.";
RN [2]
RP SEQUENCE FROM N.A.
RA VIK D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DDJ databases.
DR EMBL: L17428; AAB60695.1; JOINED.
DR EMBL: L17429; AAB60695.1; JOINED.
DR EMBL: L17418; AAB60695.1; JOINED.
DR EMBL: L17430; AAB60695.1; JOINED.
DR EMBL: L17392; AAB60695.1; JOINED.
DR EMBL: L17394; AAB60695.1; JOINED.
DR EMBL: L17396; AAB60695.1; JOINED.
DR EMBL: L17398; AAB60695.1; JOINED.
DR EMBL: L17400; AAB60695.1; JOINED.
DR EMBL: L17409; AAB60695.1; JOINED.
DR EMBL: L17408; AAB60695.1; JOINED.
DR EMBL: L17407; AAB60695.1; JOINED.
DR EMBL: L17406; AAB60695.1; JOINED.
DR EMBL: L17405; AAB60695.1; JOINED.
DR EMBL: L17403; AAB60695.1; JOINED.
DR EMBL: L17402; AAB60695.1; JOINED.
DR EMBL: L17401; AAB60695.1; JOINED.
DR EMBL: L17419; AAB60695.1; JOINED.
DR EMBL: L17417; AAB60695.1; JOINED.
DR EMBL: L17416; AAB60695.1; JOINED.
DR EMBL: L17415; AAB60695.1; JOINED.
DR EMBL: L17414; AAB60695.1; JOINED.
DR EMBL: L17413; AAB60695.1; JOINED.
DR EMBL: L17412; AAB60695.1; JOINED.
DR EMBL: L17411; AAB60695.1; JOINED.
DR EMBL: L17410; AAB60695.1; JOINED.
DR EMBL: L17427; AAB60695.1; JOINED.
DR EMBL: L17426; AAB60695.1; JOINED.
DR EMBL: L17425; AAB60695.1; JOINED.
DR EMBL: L17424; AAB60695.1; JOINED.
DR EMBL: L17423; AAB60695.1; JOINED.
DR EMBL: L17422; AAB60695.1; JOINED.
DR EMBL: L17421; AAB60695.1; JOINED.
DR EMBL: L17420; AAB60695.1; JOINED.
DR EMBL: L17399; AAB60695.1; JOINED.
DR EMBL: L17397; AAB60695.1; JOINED.
DR EMBL: L17395; AAB60695.1; JOINED.
DR EMBL: L17393; AAB60695.1; JOINED.
DR EMBL: L17391; AAB60695.1; JOINED.
DR EMBL: L17390; AAB60695.1; JOINED.
DR PIR: I73012; I73012.
DR HSSP: P17827; IGRN.
DR GO: GO:0004872; F:receptor activity; IEA.
DR InterPro: IPR000834; Peptidase M14.
DR InterPro: IPR001424; SOD CU ZN.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; Sushi_37.
DR SMART: SM00032; CCP_37.
DR PROSITE: PS00133; CARBOXYEPT_ZN_2; UNKNOWN_3.
DR PROSITE: PS00087; SOD CU ZN_1; UNKNOWN_1.
DR PROSITE: PS50923; SUSHT_37.
SQ SEQUENCE 2489 AA; 272846 MW; CER11B53F2B4FAF6 CRC64;
Query Match 9.1%; Score 1550; DB 2; Length 2489;
Best Local Similarity 20.8%; Pred. No. 1.9e-79;
Matches 676; Conservative 335; Mismatches 1031; Indels 1204; Gaps 141;
Qy 58 KTSVLTGVGVSGQHNNCPDPGIERGK-RLGSPFRIGSSVQVFPNCNGYLDQSKRTTCKK 116
Db 88 KNSVWTGAKDRCRRKSCRNPPDPVNGVHVIKGIQGSQIKYSCITGRYLIGSSSATCI- 146

```



```

Oy 117 VSDMFAASDHPRVCARACDAHLRGPSGITSPNFQIYDNNAHCWIIIT-ALNPSKYI 175
Db 147 ISGOTVIMDNETPICDRIPC-----GLPPTITNGDFTSTRENHYSVYTRCNPSGG 201
Oy 176 KLAEEFDEBERGYDITLVGDGODGDDQKTVLYMSQNAOSDHPHPSRHPESMGDIWRQ 235
Db 202 RKVEE-----LVGEPS-----LYCTSN-----DDQV-IV-- 225
Oy 236 KMYULEICRODISSDARSGSVKSPKTSNAVELVAPGTEIOGSGCDP---GIPAYGR 291
Db 226 -----SGPAPQCIIP-----NKCTPNNVENGILVSDNR 253
Oy 292 EGRFHHGDTLKFEQOPAFELVQKAITCOKNNOMSAKPGCVSPCFNFTSPSGVULSP 351
Db 254 --SLPSLNEVVEFRQCPGFVKGPFRVYKQALNKKEPELPSCSRVC-----QPPDVLYAE 307
Oy 352 NYRPNYGNHLHCWMLILARPESRIHLAFNDIDVEPODFVLVYKQATAEAVLGTFSGNQ 411
Db 308 RTQBDKNF-----SPQGEVYS-----CEGYD--LRGAASMRCTPQDMS--- 347
Oy 412 LPSSITSSGHVAREFOTDHTGKRGFNITTTFRANECPP--PGVPVNGK--RFGDSIQG 469
Db 348 -PAA-----PTCEVKSDDPMQQLNGRVLFPVNLQIG 379
Oy 470 SSISFLCDEGFLTGQSEITICVLKEGSVWNSAVLRCEAPCGHILSPSGTILSPGMG 529
Db 380 AKVPFVCEGF--QLKSSASYCVLAGMESLWSSPVCE-----QIFCESPPVILNG-- 430
Oy 530 FYKDALSCAWIEAQPGYPIKITEPRFTEVNYD-----TLFVR 568
Db 431 -----RHTKRELEV--PFGKAVNYTCDDPHDRGTSPLIGESTICTSPQ 475
Oy 569 DGRYYS--APLIGYHGTQVPOFLISTSNLYLFSTDKSHD--IGQLAYE----- 617
Db 476 GNGWSSPAPRCGLIGHCOAP-----DHFLFAKIKQTNASDPPIGTSLTECRPEYGR 529
Oy 618 ---TITLOSHCLD-----PGIPVNGQRG--NDPVGALVMTFSCDSG 655
Db 530 RPSFIT-----CLDNLWVSPKDYCKRKSCTPPDPVNGMVHTDIOVGRINYSCTG 584
Oy 656 YTLSDGEPLCEPN---FOWSRALPSCBAL--CGEFTIGSSGTTILSPGPDPYNNLNTW 711
Db 585 HRLGHSSAECLISGNAHAWSTKPIQRIPOG----- 617
Oy 712 IIFSHGKGVFTFHTFLESBGHDYLLITENGSTOPLRLQTLGSRLPAPISAGLYXNETA 771
Db 618 -----LPTIANG----- 625
Oy 772 QVRPIISDF--SMSYSGFNITFSEYDLCEPEBEVPAYSIRKGLQGVGTLLFSCFPG-- 827
Db 626 ---DITSTNRENPH-----YGSVYTRCNPSGG 651
Oy 828 ---YRLBSTARITCIGRRR--LWSSPLPRCAEGCNSVTGTGTLSPN--FPVNYNN 880
Db 652 RKVELVGEPSIYCTSNDDQVIGMSGAPQCI-----XPNCTPNNENG 696
Oy 881 HECIYSIQTOPKGIQLKARAFELSEGDVLKYDGNNSARLLGVFSHSEMGVTLNSTS 940
Db 697 -----ILVSDNRS-----LFSINBY-- 711
Oy 941 GSWLWDFITDAENTSKPELHFSSEFLICKEDPGTPKPKGYKHVBGHAAGSSVSFSCDPG 1000
Db 712 -----VEFRQCPG 719
Oy 1001 YSLGSEELICLGERRTMDRLPTCAVCEGTVAGEVSGQLSPGYAPAEHNLACIWT 1060
Db 720 FVMGCPRRVKQALNK--WEPELPSCSRVC-----QPP-- 751
Oy 1061 IEAEGCTIGLHFLVFTBEVADVLRIMWDGVEVSGVLLKELSGPALPRDLHSTNSVVLQ 1120
Db 752 -----DVLH----- 755

```

```

Oy 1121 FSTDFTSKOGFALIQFSVSTATSCNDPILPONGSRSDSWBAGDSTVFOCDPGYALOGSA 1180
Db 756 -----AERTORDDNNSPQGEVYYSCEPGDILGAA 786
Oy 1181 EISCVKIENRPFWOPSPPTIAPCGDLTP--SGVILSPNYEPBPYKBCMKVTVSP 1238
Db 787 SMRCTPOGD--WSPAFTCEVKSDDPMQQLNGRVLFP----- 823
Oy 1239 DYVALVFNIFNIEPGDPLFIHYDGRSLPLIGSPFGSOLPGRIESSNSLFLA----- 1293
Db 824 -----VNIQLGAKVDVCD-----EGFQDKG--SSASTCVLAGMESL 858
Oy 1294 FRSDASYNAGFVVDYENPRESCFDPGSIKNGTRVSDKL--GSSVYYTC-- 1344
Db 859 WNSSVVPCEQIF-----CSPSPVLPNGHCKPLIEVFPFGTVYTCDDPHDRG 907
Oy 1345 GGYEVEGTSTLSCILGDBGKPVWNNPPPVCTAPCGQVGSDDVILSPNYQNTSQIC 1404
Db 908 TSFPLIGESTIRCTSDPQNGWSSPAPRC-----GIL-- 940
Oy 1405 LYFTVVKDYVREGQAFHTHALNDVVEVDHSHQ--HSRLLSLSGSHGESLPLATSN 1462
Db 941 -----GHQADHFLFAKIKQTNASDPPIGT-- 968
Oy 1463 QVLKIFSAKGLAPARGF-----HFVYQAVPR--TSATQSSVPEPRYK--RLGSDPSVG 1513
Db 969 ---LKYECREBYRPRPSITCLDNLWSS--PKDYCKRKSCTPPDPVNGMVHTDIOVG 1024
Oy 1514 AIVARECNSGVALQGSPEIECLPVPGALAOVNSAPTCV--VPCGGLTERRGTTILSPGF 1572
Db 1025 SRINYSCTGHLRILGHSSACI--LSGNAHAWSTKPIQRIPOGLPPTIANGDPISTNRE 1083
Oy 1573 EPIYNSLNCWAKIVPEAGAIQIOVYSFVTEQWMDSELEVDDGADNTYMLGSPGTTVPA 1632
Db 1084 NFHYGS-----VWYTRCNPSGGKREVE-----LVGEPS-- 1112
Oy 1633 LNSTSNQYLHFSDISVBAAGFHELYKTVGSSCEBPAPVSPNGVKTGER--YLVNDV 1689
Db 1113 -IYCTSNDDQVIGMSGAP-----QCIXPKCTIPPNV--ENGILVSDNRSLSINBY 1161
Oy 1690 VSPQCEBRYALQGHAIHISCMPGTVRWVNPPLICIAOCGGTVEBMEGVILSPGPNGVPS 1749
Db 1162 VEFRQCPGFWKGRVRVCKQ--ALNKWPELPSCSRVC----- 1197
Oy 1750 NMDCSKIALPVGRGAIHQPLNSTEPNHYIETERNPYETSRMGRFSGSELPSLST 1809
Db 1198 -----QPPDVL-- 1204
Oy 1810 SHETTYFHSQONRPGFLLEYQAYELQCEPDEPPANGIVRGAGVNGQSYTFECLPG 1869
Db 1205 ---HABRTOR-----DKDNF-----SPQGEVYYSCEPG 1229
Oy 1870 YOLTGHBVLTCOHTNENMDHPLPKCEVPCGNTSS--NGTVYSPGFPSPSSQDCW 1927
Db 1230 YDLRGAASWMT--POGDWGPAPAPTCEVKSDDPMQQLNGRVLFP----- 1273
Oy 1928 LIIVPIGHVRLNLSLQTEBSGDFITIMGPOQTARLGVFRSMKAKTVQSSNOVL 1987
Db 1274 -VNIQLGAKV-----DFVCD--EGFQDKG--SSASTCVLAGMESL 1302
Oy 1988 KFRHDAATGFIAPAFSAVPLTK--CPRPTILPNAEVATEN--EEFNIGDIVRYRCP-- 2041
Db 1303 ---AGHESLWNSVPECEQIFCPSPPVLPNGHGTGKPLIEVFPFGAVNTTCDPHR 1354
Oy 2042 ---GFTLVGNBILITCKL-----GTYLQFEGPPICEV--HCPTNELLTSTGVILSQSY 2090
Db 1355 DRGTSFLLIGESTIRCTSDQNGW--SSPAPRCGLIGHCOA----- 1395
Oy 2091 PGSTPQOTQCSWLVREPDYINISLTVEYFISEKOYDEFELFDGSGQSPILKALSGYSA 2150
Db 1396 -----PD-----HFLFAKIKQTNASDPPIGTSLTEYECRPEYGR 1430
Oy 2151 PLIVTSSNSGVILWMS--DHAVNRKGFKIXYSAVYCSLPAPLHGFTLGQTSIQGSGSIH 2209

```

```

Db      1431 PFSITCLDNLY---WSSPKDVCCKRS-----CKTPPPVGMVAVIIDIVGSSIN 1478
Qy      2210 FGCNAGYRLVGHSMACIRHPOGYHLMSEAIPLCALSCGPPEAPKXMGV---KEYTV 2266
Db      1479 YSCTGHHILGHSSAECLISGNTAH-WSTKPPICQRIKPPPTIANDPISIRENFHY 1537
Qy      2267 GTRAVYSCSEGYH---LQAGNATAECL---DTGLMS-----NRVVP----- 2303
Db      1538 GSVVYTRCNLGRKRVFELVGEPSIYCTSDNDQVIGMSGARPOCIIPNKCTPPNENGI 1557
Qy      2304 -----OCVP-----VTC-----PDVSSIS-----VEHGR 2322
Db      1598 LVSDNRLSFSINEVEFRCOGFVWKGRPRVCKCALNKEBELPSCSRVCOPPEILHGE 1657
Qy      2323 WRLIFETOYFOAOQLMLICDPGYVYTGQVIRRCQANGWSIGDSTPICRITISC---GEL 2378
Db      1658 HTPSHQDNFSPGOEYFSCBEPGYDIRGASLHCTPQGWS--PRAPCAVKSCDPLGOL 1715
Qy      2379 PIPNGHRIGTSLV-YGATAIFSCNSGYTLVGRBECMANG--LMSGEVRCLAGHC 2434
Db      1716 ---PHGRVLPPLNLQGAKVSFVCDDEGRLKSSVSHCVLGMKSLMNNVPCVCHIFCP 1772
Qy      2435 TPEEIVNGHNGE---NYSYGSVYVQCN-----AGRLIGMSVRIQQDHH---WSG 2481
Db      1773 NPAILNGRHTGTPSGDI PYGKEISYTCDPHDPHGMFTNLIGESTICTSDPHGNGVWS 1832
Qy      2482 KTRPC-VPIITCGH-----PGNPVNLGTQGNQFNLDVYKFCNGVMAEGARS 2529
Db      1833 PARCELSTVRAGHCKTQEPFPAPFTPIPFND---EPVGTSLINECRGTYGKMFES 1888
Qy      2530 QCLASGOWMLPTCRILINCTDPEHQNSVROVHASGPHRFPQTIVSYRCHNFPYLLGT 2589
Db      1889 -CLENLWSSVEDNCRKRS CGPPEPENGWNIWTD---TQGSITVYSCNIGFRLLG 1943
Qy      2590 PVLSCQGDG---TWDRRPOCLIVSGHPSGSPHSGSDSYR-----VGAIVKXSC- 2638
Db      1944 PSTTCLVSGNNVYTDKAPICEIIS---EPPTISNGPYNNRTSPHNGVTVVYQCH 1999
Qy      2639 -----IGKRLVGNSTRMCGLDGHWTSGLPHCSGTSYVCGDPGIPAHGIRLGD 2687
Db      2000 TGPDEQLFELVGBRSIY--CTSKMDQVGVWSPPRCISTN--KCLAPRV-ENALRVG 2054
Qy      2688 S---FDGTYMRSCEAGHILRGSSERTQANGSWSGOPECVATCGNPGTPSNARVVF 2744
Db      2055 NRSFSLTELIVRRCQGFVWVGSHTVQCCTNGRMGPKLPHCSRV-COPPEILHGEHTL 2113
Qy      2745 S--DGLVFSSSIYVECEGYATGLSRHCSVNGTWTGSDPECLVINCGD--PGIPANGL 2800
Db      2114 SHQDNFSPGOEVYVSCBPSYDLAAGASLHCTPOGMSPEAPRCTVKSCDPLGOLPHGRV 2173
Qy      2801 RLGNDRRYNKTVTYQCVPGYMESHVVS-VLSTCRDRTNGTKPYCKALMCKPPLIPN 2858
Db      2174 LLEPLNLQGAKVSFVCDDEGRLKGRASHSCLVLAGMK-ALMNSVPCBQIFCPNPAIIN 2232
Qy      2859 GKTVGS---DFWNGSSVTVAC-----LEGQLSLPAVFTGEGNGSTGELPQC- 2903
Db      2233 GRKTGTPFDPIPYGKEISYACDTHPRGKMTFNITG--ESSIRCTSDRQNGVWSSPAACE 2291
Qy      2904 --FPEVCGPGEVRSRGRREDRGFSY--RSSVSFSCPLVLVGSRRRCQSDGTWSGTQ 2958
Db      2292 LSVPAACRPBKILONGHYIGHNSLPLPMITISYICDPEGLVKGKFICTOGIWSQID 2351
Qy      2959 PSCIDTTLTTCADPGVPOGIONN---SGCYOVGSIVLFRCKGKYLLOGSTRTCLPINT 3015
Db      2352 HYCKE---VNCSPF-LFMNGISKELMKKVHYGVYTLKCBGYTLGSSPWSQCADDR 2407
Qy      3016 WSGTTP 3021
Db      2408 WD--PP 2411

```

RESULT 9

```

07LCS3
ID 07LCS3 PRELIMINARY; PRT, 3494 AA.
AC 07LCS3;
DT 05-JUL-2004 (TREMblrel, 27, Created)
DE 05-JUL-2004 (TREMblrel, 27, Last sequence update)
DE 05-JUL-2004 (TREMblrel, 27, Last annotation update)
DE Intrinsic factor-vitamin B12 receptor (Fragment).
CN Name-CUBN;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SOURCE FROM N.A.
RA Antinoff M., Brady S., Verroust P.J., Moestrup S.K., Krahe R.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
DR EMBL; AF243129; AAK61830.1; -.
DR EMBL; AF243083; AAK61830.1; JOINED.
DR EMBL; AF243084; AAK61830.1; JOINED.
DR EMBL; AF243085; AAK61830.1; JOINED.
DR EMBL; AF243086; AAK61830.1; JOINED.
DR EMBL; AF243087; AAK61830.1; JOINED.
DR EMBL; AF243088; AAK61830.1; JOINED.
DR EMBL; AF243089; AAK61830.1; JOINED.
DR EMBL; AF243090; AAK61830.1; JOINED.
DR EMBL; AF243091; AAK61830.1; JOINED.
DR EMBL; AF243092; AAK61830.1; JOINED.
DR EMBL; AF243093; AAK61830.1; JOINED.
DR EMBL; AF243094; AAK61830.1; JOINED.
DR EMBL; AF243095; AAK61830.1; JOINED.
DR EMBL; AF243096; AAK61830.1; JOINED.
DR EMBL; AF243097; AAK61830.1; JOINED.
DR EMBL; AF243098; AAK61830.1; JOINED.
DR EMBL; AF243099; AAK61830.1; JOINED.
DR EMBL; AF243100; AAK61830.1; JOINED.
DR EMBL; AF243101; AAK61830.1; JOINED.
DR EMBL; AF243102; AAK61830.1; JOINED.
DR EMBL; AF243103; AAK61830.1; JOINED.
DR EMBL; AF243104; AAK61830.1; JOINED.
DR EMBL; AF243105; AAK61830.1; JOINED.
DR EMBL; AF243106; AAK61830.1; JOINED.
DR EMBL; AF243107; AAK61830.1; JOINED.
DR EMBL; AF243108; AAK61830.1; JOINED.
DR EMBL; AF243109; AAK61830.1; JOINED.
DR EMBL; AF243110; AAK61830.1; JOINED.
DR EMBL; AF243111; AAK61830.1; JOINED.
DR EMBL; AF243112; AAK61830.1; JOINED.
DR EMBL; AF243113; AAK61830.1; JOINED.
DR EMBL; AF243114; AAK61830.1; JOINED.
DR EMBL; AF243115; AAK61830.1; JOINED.
DR EMBL; AF243116; AAK61830.1; JOINED.
DR EMBL; AF243117; AAK61830.1; JOINED.
DR EMBL; AF243118; AAK61830.1; JOINED.
DR EMBL; AF243119; AAK61830.1; JOINED.
DR EMBL; AF243120; AAK61830.1; JOINED.
DR EMBL; AF243121; AAK61830.1; JOINED.
DR EMBL; AF243122; AAK61830.1; JOINED.
DR EMBL; AF243123; AAK61830.1; JOINED.
DR EMBL; AF243124; AAK61830.1; JOINED.
DR EMBL; AF243125; AAK61830.1; JOINED.
DR EMBL; AF243126; AAK61830.1; JOINED.
DR EMBL; AF243127; AAK61830.1; JOINED.
DR EMBL; AF243128; AAK61830.1; JOINED.
DR GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR000152; Aex_hydroxyl_3.
DR InterPro; IPR000859; CUB.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00431; CUB; 27.

```

DR Pfam; PF00008; EGF, 4.
 DR Pfam; PF07645; EGF CA; 3.
 DR SMART; SM00042; CUE; 27.
 DR SMART; SM00181; EGF; 8.
 DR SMART; SM00179; EGF CA; 7.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS01180; CUB; 27.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01166; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 5.
 DR PROSITE; PS01187; EGF CA; 3.
 DR EGF-like domain; Receptor.
 FT NON TER 1
 SEQUENCE 3494 AA; 384281 MW; 0A811043985DB39 CRC64;

Query Match 8.78; Score 1473; DB 2; Length 3494;
 Best Local Similarity 21.28; Pred. No. 8e-75;
 Matches 795; Conservative 416; Mismatches 1184; Indels 1356; Gaps 174;

113 TCCKVSDMPAA-----WSDHPRVGR--RMCDALHRLPSGIITSPNPIQVDDNNAHCWMI 165
 317 TCVDGVVSFSCETRLMTG--ALCQVPOQVCGESLSGINSFSYRSPDVGVHDVNCFWV 374
 166 ITALNPCKVILKAEFEFLER---GYDTLVGDGQDQKTVLYMSQNAQSDSPHTPG 221
 375 IKT-EMGKVLRTFTFRLESMNDCRHEFLQV---YDGDSSAFQIGRFGSSLRH--- 426
 222 SRIPESMSGDIWROKMTVLEICRDISSDARSQSVKRSPTKSNALVAPETEIEQSCG 281
 427 -----ELLSSD-----NALYFLYSEHLRNG--- 447
 282 DPGIPAYGRRGSRFHHDITLKECORAFELVQKAITCKONNWSAKKPCVNSCFPNF 341
 448 -----RGFTVR-----WETQDP---ECGAIL 465
 342 TSPSGVLSPNYPEDYGNHLHCWLLIARPESRILHAFNDIDVPR---QFDPLVIKGA 397
 466 TGPRTKISKPGYPCNYPGRDCWMIWVTSPLVLTFTFGTISLEHDDCKMDVLEIDGR 525
 398 TAAEPLVIGTSGNOLPSSITSSGHVAREFQTDHSTGRKGNITFTFRHN-ECPPDGPV 456
 526 LYODPRLGKCTTFSVPLQTTGPFAIHFHSDQISDQGFHITLYLTPSPDLRCGNVTD 585
 457 VNGRRF-----GDSLQG--SSISFLCD-----EGF 480
 586 PEGELFLPELSGPFTHTRCVYMMKOPQEGIQINFHVELQCCSDSSQNYIEVRDETL 645
 481 LG--TQSGEITTCVLKEGSVVV-----NSAVLRG--EAPCGHLLTSPGTTLSGMP 528
 646 LGKVCNGGTISHIKSINSTNSWIRKIDASVEKASFRAVYQVACGDELTG--EGVIRSFPR 704
 529 GFYKDALSCAMVIEAOPGYPIKITFDRPK---TEVNYDLEVRDGTYSAPLIGVYGT 584
 705 NVYPERGECRWHTIHQPOSQVILLNFVFEIGSSAHCEITDYVEIGSSIIISPENKXCGT 764
 585 QVPEFLSTSNLYLLFSTDKSHSDIGFQLRVETITLQSDHCLDPGLPVGQRKGNDFV 644
 765 DIBEFITSVNPLVTVTKSSSTENHGFMAKF----- 796
 645 GALTTFSCDGSYTLDSDEPLECEBPNQWRSRALPSCAL--CGGFIQSSGTTLSGPFDFY 703
 797 -----SADLACGEILTESTGIIQSGHNVY 823
 704 PNNLNTWIIETSHGKGVFTFHTFLE---SGHDYLLITENGSTFQPLRQLTGSRLPA 759
 824 PHGINCNTWIIIVQPNHLIHLMFETFLHFNCTNDYLEVYDTSETS--LGRYCGKSIIP 882
 760 PISAGLYGNFTAQVRFISDPSMSYEGNITTSSEVDLAPCEBPVPAISIRKGLQFGVGT 819
 883 SLTSS--GN--SLMLVFYTDSDLAEGFLINEAI----- 913
 820 LTFSCFPGRLEGARITCLGRRRLWSSPLPRCVAECGNSVTGTGTTLSPPNPNVYNN 879

914 -----SAATACTL-----QDYTDLCTFTSPNPNPNY 941
 880 NHECTYSIQTPKGIOLKARAFELSG-----DVLKATVDGNNSAKRLIGVSHSEM 933
 942 NHECTYRITVTRQOLAVHFTNFSLEAIGNYYTDFLEIRIDGGEKPELGIIF--YGSNLP 1000
 934 VTINSTSSSLMLPITDAENTYSGFELHFFSFEIKECDGTGPKFGKYVDEGHFAGSSV 993
 1001 PTIISHNKLMLKFKSQIDTRSGFSAYWD-----GSSTCGGLNLTSSG 1045
 994 SFGCDPGY-----SLRGSBEL-----CLSGERRTDRP----- 1022
 1046 TF-ISPVPMPYHSSRCYMWLSSHSGSAFELEKDFLHHNCTLDLYAVVDGPFSSNS 1104
 1023 -----LPTCAECGTV-----RG-----EVSQVLS 1044
 1105 HLLTOLCGDEKPLIRSSGSMFIKLTDEGQGRKAEYROTCEENVIVNQTGILBS 1164
 1045 PGYAPAYEHLNLCMTTEAAGCTIGLHPLVFTDEVH-----DVLRIWDGPVSSGLK 1099
 1165 IGTNPYSEKHQCNWITRAITGNTVNTFLAFDLEH-HINCSITDYLELYDGPROMG--- 1219
 1100 ELSGPALPKDLHSTFNSVVLQFSTDEFTSKQ--GRAIQSV-----STAT--SCNDGP 1150
 1220 RYGVVDLPPRGSTTSSKQLVLLTDGVRREKFGQWQMFYVCGGSGELSGATGFSFGPF 1279
 1151 QNSRSGDS--WEAGDSTVPCDPGYALQ-----GSAEISCVKI 1187
 1280 NRYPNKCEIWM-----YIRTDGSSIQITIHDFVEYHRCNFDVLEIYVGDPFHSPI 1333
 1188 -----EN-----RF-----FMOPSPPTCIAPCGGDLTGPS 1212
 1334 AOLCTORSPEPNQOVSTGNELAIRFKTDLISNGRCNMAWQ---AVTGCGGIFAPPS 1389
 1213 GVILSPNYPPEYRPGKECDWKVTVSPDYVALVFNIENLPEGYDFLHIYDGRDS--LSPLI 1271
 1390 GEIHSPPVPSPYRSTNDSCAVIVDRYHRLVLPFTDLEPQSCIMAYDGLSTMSRLA 1449
 1272 GSFTYGSQLPGRISSSNSLFLARSDASVSNAGFVIDYTNPRESC-----FD--- 1319
 1450 RTGREGALAPRIVSSGNSLFLARFOSGPRNRRFRAQF---RQACGHILITSPDFVSS 1505
 1320 ---PGSIKNGTR---VGSDLKGSVTVYCHGVEYEGTJLSC---TL--GPDGKPV 1366
 1506 PREPANTPNQNCSWIIQAOPPLNHTLSFTH--FELERSTICARDEVELLDGHEAPL 1563
 1367 -----WNNRPVCT-----APCGGYVSGDVVLS 1392
 1564 RGRYCGTDMHPHITTSFSSALTLRFVSDSSISAGFHTTVTAASACGTFYMAEGIFNSP 1623
 1393 NYQONTYSQICLY-FVTVPKDYVVFQPAFFHTAL-----NDVVEVHDSHQSRLL 1444
 1624 GYDPIYEPNVECWNIISSPGNRL--QLSFIFOLEDSODCSRDFEIREGNA---T 1675
 1445 SSLSGHSTGSLPLATSNQV---LKFSAKGLAPARGHFVYQAV----- 1486
 1676 GHLVGRICGNSFPLNTSSIVGHTLWRFISDGSSTGFOATFMKIFGNDNIVTGKGA 1735
 1487 -----PRTSATQ-----CSSVPEPRY----- 1502
 1736 SPFWPENYPHNSNYQMTVVNNASHVHGRILEMDIEIQCYDXKLRIYDGRSHARLIG 1795
 1503 ---GKRLGDPFVGAIVFPCNSGVALQSPFEIECLVPGALAQM--NVAAPTCVVP--- 1554
 1796 AYCQTOTESFSSGNSLTFHFYDSSISGR-----GFLLEWFAVDAPDGLPTIAP 1846
 1555 ---GKRLGTERGT-----IISGRPEPLNLCNVKIVPEGAGIQOVVSPFTEQN--- 1605
 1847 GACGGFL--RTGAPVFLFSPGMPDYSNNVDCITMLIQADST--VELNITSLDIESRRT 1903
 1606 -WDSLEVFADADNTVTMLGFSGTVPALINSTNQLYLFHYSDISAAGFHELYKTVG 1664
 1904 AYDSLVIRDDGNLQAQGLAVLCGREIRPGIRISGEYWFIRFTSDSSVTRAGFNASPH--- 1960

QY 1665 LSSCEPAVNSGVTKGERLYLVNDVVSFOC-----EFGYALQGH-----AHISCM 1709
 DB 1961 -KSCGGYVLAHADGIIITSPKPIETTPYNSLNCNHWLVQSGLTIAVFEOPFOIPNGDSGSCN 2019
 QY 1710 PG---TVRRMN---YPPPL-----CIAQ----- 1726
 DB 2020 QGDYVLRNPGDIPYSPPIQPCGNGHFGGSHASSTLFTSDNQMFQVPISDHNSGQGHKI 2079
 QY 1727 -----CGGTUV-----BEMGVILSPGPGNYPSNNMDCSKIALPVGFAGHIOF---LN 1771
 DB 2080 KYEAKSLACGGNVTYHDAISAGYVTSPPHFNHYPPHADCIWILAAPETRIQLQFEDRFD 2139
 QY 1772 FSTERN---HDYIEIRNGPYETSRMMGRFSGSELPSTLSTHETTYVPHSHQNRGPKF 1829
 DB 2140 IEVTNCTSNVLELADGVSDAPILSKTCGTSLSBSQMSGCEWVWYLRFRSDNSFTHVGF 2199
 QY 1830 LEQAYEIQECDEPPFPFANGIVRGAGY-----NVGOSVTECLPGYOLT----- 1873
 DB 2200 AK---YSIAOCGRVPGQSGVYESIGHPTLPYRDNLFCEMHLQGLSGHYLTISFEDFNLQ 2256
 QY 1874 ---GHPVLTCQHGTRNMD-----HPLP----- 1893
 DB 2257 NSSG-----CEKDVYEIWDNHTSGNILGRYCGNTIPDSIDTSNTAVRREVTDGSVYASG 2311
 QY 1894 -----KCEVPCCGNITSSNGTVVSPGFSPYSSSDCWMILITVPIGHGYRL---NLSLQ 1945
 DB 2312 FRLRFESSMEBCGDLQOSIGTFTSPVNPNNPHGRICEMRITAPESRRITLIMFNLRLA 2371
 QY 1946 TEBS---GDFITMNGPOOTAPRLGVFTFRSM-AKTVOSSSNOVLKFRHDAAT---GGIFA 2000
 DB 2372 THPCNNHEVTVFNGIISNSFOLEKCSVAVNSNEIKSGNTMKVIEFTFGDSRPYCG-FT 2430
 QY 2001 IAFSAVPLTKCPPTLIPNAEVVTENEEFNIGDIVRCLPGFTLVGN----- 2048
 DB 2431 ASYSSSDAVCCGS--LPN---TPBGNFT---SPGIDGVANYSRNINCEWTLS 2475
 QY 2049 -----ELTCKLGTYLQFEGP-----PRICE 2069
 DB 2476 NPNQGNSSISIHEDFYLBESHODCOFEVLEFRVVG---DADGPIMLMRLCGPSKPTLVLVP 2532
 QY 2070 -----VHCPTNELL-----TD-----STGVILISQSPSSYPOFQICSMVLVY 2106
 DB 2533 YSQWVHFTVNERVEHIGFAKYSFTDCGGIQTIGDGGVITSPVPAVYSLTHCSMLTFA 2592
 QY 2107 EDPYNISLTVYFLSEKO---YDFEIPDPSGSGPLKALSGNYSAPLIVTSSNSYV 2162
 DB 2593 POGHTITLTFSDPDIEBHHTCAMDSVTVNRNGSPESPITIGQYCGN-SNRTTIGSSNOLV 2651
 QY 2163 LRASSDHAVNRKGFKIRISAPYCSLRAPLHGFTLQJOTSTOPEGSIHFGCNAGYRLVGH 2222
 DB 2652 VTNSDHSLOGGGP-----YATWNTQT-----LCCGG-----IFHS 2682
 QY 2223 MAICTRHPQGYHLMSEAIPLCALSS-CGLPEAPKN-GWFEKGYTVGTAKAVYSCSEGYHL 2280
 DB 2683 DNGTISPH---WPNFNPENSRCSWTALTTHSKHLEIFDNNFLIPSDGOCQCSFYKV 2738
 QY 2281 QAGAEKTAAC-LDTGILMSNRNVPPOCVPTCDVSSISVEHGRWMLIFETQYQFOAQLML 2339
 DB 2739 WAGTEVVDKALLATGC---GNVAPG---PVITPSNFTAV-----FQSQA- 2778
 QY 2340 ICDPGYVYTGQRYIRCOANGKMSLG-----DSPTCRITISGCELPPIPPNGHRIG 2388
 DB 2779 ---PAGFSASFVSRKGSNFTGPGYIISPNYPKOYDNNMNTYV-----LEAN----- 2824
 QY 2389 TLSVYGATAI-FSCNSGYTLVGRVRECMANGLMGSEVRCLAGH-----CG--T 2435
 DB 2825 PLSVVLLTFVSPHLEARSATVS---CVNDG-----VHTRKGSVWSTPATVYCGDEM 2874
 QY 2436 PEPI-VNGHINEBNYSRGSVYVQCNAGRLLIGMSVRIQQODHMSGKTPCVPITTCGP 2494
 DB 2875 PAPTITAGVLLNLFYSNEQITDF---GPK---FSYRI-----ISCGGV 2911

QY 2495 GNPVNGLTQGNQPNLNDVVKVCNPGY-MAEGAARSQCLASGQWSDMLPTCRINCTDPC 2553
 DB 2912 FNFSSGI-----ITSPAYSVADVPNDHCHLYTITVSD----- 2943
 QY 2554 HOENSVROYHAGSPHAFSGTIVYSRCHNGFYLLGTVPVLSGQDGTWDRPPOCLLVSCG 2613
 DB 2944 ---DKVIELEKFS-----DEDVVPSTSCSHDYAL-----YOGANTSD---PLGKFCG 2985
 QY 2614 HPGSPHQSOMSG-----DSYTVGAVVRSCIGKRTLVNSTRMGSLDHMGCSLPH 2664
 DB 2986 SK-RPENVASNNMSMLVETKIDSFOYAKMKRFS--RQTL--GPOQCG--GYLTG----- 3034
 QY 2665 CSGTSVGVGDGPIPAHGIRLDSFDPG-----TVNRFSCGAGVLR- 2706
 DB 3035 ---SNNTPASPDSDSNGM---YDKNLNCWIIIAPVNKVTHLTFTFALFAASTQR 3085
 QY 2707 -----GSSEPTCOANGSMGSGQEGCVISGNGCTSPNANVVSDDLVESSSIY 2756
 DB 3086 CLYDYVKLYDGSSEANALIA-GTEFCSTVPAPFISGNEPLT-----VOFISDLTL 3133
 QY 2757 ECEGYAT-GLLSRHC--SVNGTWG--SDPECLVINCGDPGIPANGRLGNDFRYNK 2810
 DB 3134 E-REGFNATYTIMDWPCCGTVAWTQNISSP-----NSSDPVPFS----- 3175
 QY 2811 TVTYQCVGYMMESHRSVLSCTKDRTWNGTKPVCKALMCKPPLIPNGKVVGSDFMWS 2870
 DB 3176 -----ICTWIDSPPH--QQVKIT-----VVAL 3196
 QY 2871 SVT-YACLEGYQLSPAVFTCEGNS---WTGELPQCFPFECGD-----PGVPSR 2916
 DB 3197 QLTSDDCQNY---LQLODSQGHONBSFOCGRNASAVPFFYMSGTAMVIFSGVNR 3253
 QY 2917 GRREDGYSY-SSVSFSGHPPLVVSPPRFQSGDGTWGSTOCSIDPTLTTCADBPV 2975
 DB 3254 NSR--MSFTYQIACNDRDYHAKFGLNSP-----GMPDNYDNDDXCTVTLA---P 3299
 QY 2976 Q-----FGIONSGQVQVGSYVLFRCQXYL-----LOGSTTTCLPN- 3013
 DB 3300 QNHITSLFPHSLGIENSV-----CRNDFLEVRNGSNSNSPLGKYCGTLLPNP 3348
 QY 3014 -----LTWGTTPDCVPHHCHROPETPTANVAGLDPMSGY 3049
 DB 3349 VFSQNNELYLRFKSDSVTSRGEYIINTSPSGC-----GGTLYGDGCSFTSP--GY 3398
 QY 3050 -----TLITAPR-----ASPS----- 3061
 DB 3399 PGTYPNNTYCEWVLVADAGRLVTINFYFISIDDPGDCVQNYLTLYDGPMASSPSSGPGY 3458
 QY 3062 ---RVAPSTAPAR---RMAAGQARPSAMR 3085
 DB 3459 GDTSIAPFVASSNOVFIFKHADVAKRRPSAFR 3489
 RESULT 10
 ID 060494 PRELIMINARY; PRT; 3623 AA.
 AC 060494; 096RUG;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Intrinsic factor-Bi2 receptor precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97476251; PubMed=9334227;
 RA Birn H., Verroust P.J., Nexo E., Hager H., Jacobsen C.,
 RA Christensen E.J., Moestrup S.K.;
 RT "Characterization of an epithelial approximately 460-kDa protein that
 RT facilitates endocytosis of intrinsic factor-vitamin B12 and binds
 RT receptor-associated protein.";

RL J. Biol. Chem. 272:26497-26504 (1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98241400; PubMed=9572993;

RA Kozureki R., Kristiansen M., Slietcaroglu A., Hansen C., Jacobsen C.,
Tomerup N., Verrouet P.J., Moestrup S.K.,
RT "The human intrinsic factor-vitamin B12 receptor, cubilin: molecular
characterization and chromosomal mapping of the gene to 10p within the
RT autosomal recessive megaloblastic anemia (MGAI) region."

RL Blood 91:3593-3600 (1998).

DR EMBL; AF034611; AAC82612.1; -

DR PIR; T09456; T09456.

DR HSSP; Q9UCV4; INZ1.

DR Genew; HGNC:2548; CUBN.

DR GO; GO:0016030; C:membrane; TAS.

DR GO; GO:0004872; F:receptor activity; TAS.

DR GO; GO:0005215; P:transporter activity; TAS.

DR GO; GO:0007582; P:physiological process; TAS.

DR GO; GO:0015889; P:vitamin B12 transport; TAS.

DR InterPro; IPR000152; Asx_hydroxyl_S.

DR InterPro; IPR000859; CUB.

DR InterPro; IPR00742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR006209; EGF_like.

DR InterPro; IPR003030; Grow_fac_recept.

DR Pfam; PF00431; CUB; 27.

DR Pfam; PF00008; EGF; 4.

DR Pfam; PF07645; EGF_Ca; 3.

DR SMART; SM00042; CUB; 27.

DR SMART; SM00179; EGF_Ca; 3.

DR PROSITE; PS00010; ASX_HYDROXYL; 4.

DR PROSITE; PS01180; CUB; 27.

DR PROSITE; PS00022; EGF_1; 4.

DR PROSITE; PS01186; EGF_2; 2.

DR PROSITE; PS00026; EGF_3; 5.

DR PROSITE; PS01187; EGF_Ca; 3.

KM EGF-like domain; Receptor; Signal.

FT SIGNAL 1 24 Potential.

FT CHAIN 25 3623 Intrinsic factor-B12 receptor.

SQ SEQUENCE 3623 AA; 398993 MW; A91A7CEA53488651 CRC64;

Query Match 8.7%; Score 1473; DB 2; Length 3623;
Best Local Similarity 21.2%; Pred. No. 8.4e-75;
Matches 795; Conservative 416; Mismatches 1184; Indels 1356; Gaps 174;

QY 113 TCMKVSDFEA-----WSDHPRVCR--RMCDAHLRGPSSGIIITSPNPFIOYDNNACVWI 165
DB 446 TCVDGVDSFSCCECTRLMTG--ALQVPOQVCGESLSGINGSFSYRSPDVGVDVNCFWV 503
QY 166 ITANPENVITLAEEDPLER---GVDTLTVGCGGQDQDKTVLYNSQNAQCSPTPTG 221
DB 504 IKT-EMGVLLRITTFPLESMDCNPEFLQY---YDGSSSAFQLGRFGSSLPH--- 555
QY 222 SRIPESMGDIWRQKMTVLEICRDISSDARSQSVKSPKTSNVELVAPGTLEQSGC 281
DB 556 -----ELSSD-----NLYFLYSEHLRNG--- 576
QY 282 DPGIAYGRREGSFHNGDTLKPECQPAFELVGKALITCQKNQWAKKQCVESCFNF 341
DB 577 -----RGFTVR-----METQOP-----ECGIL 594
QY 342 TSPGVSLLPYRVDYGNHLCVLLARPSRHLAFNDIVP---QDPLVITDGA 397
DB 595 TGPFGISKSPGYPNYPRPDCVWIVTSPDLVTFFTGLSLHHDDCNKDVLEIRDP 654
QY 398 TAAEPLVIGTFSGNLPSSITSSGVARLEFQDTHSTGRGENITFTFRHN--ECPDGV 456
DB 655 LYQPLBLKFTCTTSVPELVLTGTPARLHHSQISIQGPHITLYLSPSLRCGAYTD 714
QY 457 VNGKRF-----GDSLQLG-SSISFLCD-----EGF 480
DB 715 PEGELFLPELSGPFTHTRQCVMMKQPOGBOIQINFTHVELQCSQSSQNYIEVRDETL 774

QY 481 LG-TOGSETTICVLKEGSVVM-----NSAVLRC--EAPCGHLTSPGTLSPGMP 528
DB 775 LGKVCNGTISHIKSTINSWIRPKIDASYEKSAFRAVYQACDELTC--EGVIRSPFP 833
QY 529 GPKYDALSCAMVIEAOGVPDKITFDPRK---TEVNYDILEVNDGTYAPLIGVYGT 584
DB 834 NVYPERGTCWTHIQPOQVILLNFVTFEIGSSAHCEYDVEIGSSIIISPENKRYCGT 893
QY 585 QVQFPLISTSNLYLLEFSTKSHSDIGFQALRYETITLQSDHCDPLGIPVNGQRGNDFV 644
DB 894 DIPSFITSVNFVLTVTVKSSSTENHGFMAKF----- 925
QY 645 GALTVPSCDGGVTLSDCEPLCEBNPQMSALPSCEAL--CGGFIQSSGTLSPGPDFY 703
DB 926 -----SABLACGEILTESTGITQSPGHNVY 952
QY 704 PNNINCTWIIETSHGKGVFTFHTFHE---SGHDYLLITENGSPQPLROLGSRIPA 759
DB 953 PHGINCTWIIIVQNNHILHLMFETPHLEFPHYNCTNDLEYVDNSETS--LGRVCGKSI 1011
QY 760 PISAGLYGNTAQVRFISDFSMSEGNITFSEYDLEPCEPEVPAVSIKGLQFGVGD 819
DB 1012 SLTSS--GN-SLMLVFYTDLDLAYEGFLIYVEAI----- 1042
QY 820 LTFSCEFGYRLGEGTARITCGRRRLWSSPLPRCVACGNSVTGTGLSPNPFVYNN 879
DB 1043 -----SATAACL-----QDYDDLCTFTSPNPNY 1070
QY 880 NHECTVSIOTOPGKGIOLKARAFELSEG-----DVLKYVDGNNNSARLLGVSHSEM 933
DB 1071 NMECTIVITRTQGLIAVHFTNFSLEAIGNVYTDLEINDGGEKSPFLGIF-YGSNLP 1129
QY 934 VTLANSTSSLMDFITDAENTSKGFELHESSEFELIKCEDGTFRKYKNDGHPAGSSV 993
DB 1130 PTIISHNKLMLFKSPQIDITRSGFAYWD-----GSTCGGGLTTSSG 1174
QY 994 SFSQDPGY-----SLRSEEL-----CLSGERTWDRP--- 1022
DB 1175 TF-ISPVPMPYTHSSCTYMWLKSHSASAFELKPDHLEHNPCTIDYLAVIDGSPSS 1233
QY 1023 -----LPTCVAECGTV-----RG-----EVSQVLS 1044
DB 1234 HLLTQLGDKRPLIRSSGDSWFKLRTDEGQGRGKAYRQTCENVVIVNQYIGLES 1293
QY 1045 PCYPAPEYHNLNCTWTAEAGCTIGLHVFPTVEYH---DYLRTWGPVSEGVLLK 1099
DB 1294 IGYPNPYSENQHCWITRATTGNTVNTYFLAFDLEH--HINCSIDYELLYDGPROMG--- 1348
QY 1100 ELGSPALPKDLHSTFNSVVLQFSDPFTSKO--GPAIQFSV-----STRT--SCNDGP 1150
DB 1349 RYCGVDLPPRGSTTSSKLVLLLDGVGREKEGQOMQVFYCGGELSGATGSFSSGFP 1408
QY 1151 QNGSRSGDS--WEAGDSTVPOCDPGYALQ-----GSAEISCVKI 1187
DB 1409 NRYPNNECIW-----YIRTPGSSIQLTIHDPDVEYHRCNPFVLEIYGDPHSPRI 1462
QY 1188 -----EN-----RF-----FWDSPPCTIAPCGDLTGPS 1212
DB 1463 AOLCTQRSPENPMQVSSSTGNELAIRFKTDLISNGRGNASWQ---AVTGCGGIFQAPS 1518
QY 1213 GVILSPYRPPYRPGKCDMKVYSPDYVALVFNIFNLEPGVDFLHIYDGRDS--LSPL 1271
DB 1519 GEIHSPPYRPPYNSNTDCSVIRDRYHRLVLFDFDLEPQDSCTIAAYGLSTMSRLA 1578
QY 1272 GSPYQQLPQRISSSSNSLFLAFRSDASVSNAGFVIDYTENPRESC-----FD----- 1319
DB 1579 RTQGRBLAMPYISSGNSLFLAFQSGPSRQNRGFAQF---RQACGHLITSSFDVSS 1634
QY 1320 ---PGSLKNGR---VGSPLKSSVTVYCHGEYVEGSTLSC---IL--GPDGKV 1366
DB 1635 PRFPANPNQNSWITIQAOPLNHTLISFTH--FELERSTTCARPFVEILDGCHDAPL 1692
QY 1367 ---WNNRPVCT-----APCGGYVGSDDGVVLS 1392

Db 1693 RGRYCTMPHPIITSSALTLRFVSDSSISAGFHTTASVSACGCTFMAGIENSP 1752
 QY 1393 NYPQVYTSQOCLY-FVTVPKDYVVFQGFAPFHTL-----NDVVEVHDSHQSRLL 1444
 Db 1753 GYDPIPPVNEBCVMNIISSPGNRL---QISFISFOLEDSODCSRPVPIREGNA-----T 1804
 QY 1445 SLSGSHTEGSLPLATSNQV---LIKFSAGLAPARGFHFVQAV----- 1486
 Db 1805 GHLVGRYCGNSPFLNYSIVGHITLMVRFISDSSGSGTFOATPMKIRGNDNIYVTHGKVA 1864
 QY 1487 -----PRTSATV-----CSSVPEPRY----- 1502
 Db 1865 SPFWEENYHNNSYQWTVNVAHSVHGRILEMDIEIIONCYDKLRIDGPSIHARLIG 1924
 QY 1503 ---GKRLGSDSVGAIVAFECNSGVALQGSPEIDCLPVPGLAQM-VNSAPTCVP--- 1554
 Db 1925 AYCGTQTESFSSSTGNSLTFHFYSDDSSISGK-----GFLEWFAVDAPDGLPTIAP 1975
 QY 1555 ---CGNLTERRG-----ILSPGFPEPYLNSLNCVWKIVPEGAGIQIQVASFVEON--- 1605
 Db 1976 GACGGFL--RTGDA.PVFLFSPGMPDSYNSRVDCTWLLQAPDST-VELNIIISLDIESHRC 2032
 QY 1606 -WDSLEVEDGADNVTVMLGFSGTTVPALNSTNQYLHPSDITSVAAGFHELYKTVG 1664
 Db 2033 AYDVLVRDGDNNILAOQLAVLCGREIFGPIRSTGEYMFIRFTSDSVTRAGFNASFH--- 2089
 QY 1665 LSSCPREBAVPSNGVKTGERYLNDVVSFOC-----EPGYALQGH-----AHISCM 1709
 Db 2090 -KSGGGLVLRDGLITSPKYPETYPNSLNCNMHVLVSGLTIAVHFQPOIENGSSCN 2148
 QY 1710 PG---TYRRNY--YPPUL-----CIAO----- 1726
 Db 2149 QGDYLVLRNGBDIYSPILGPPGGNGHFCGSHASTLETSDNQMFVQFISDHSNEGQFCK 2208
 QY 1727 -----CGGTV-----EEMEGVILSPFGPNYPSNMDCSMKIALPFGFANHOF---IN 1771
 Db 2209 KYEAKSLACGNVYIHDADAGVYTSNHPHNPPEHADCIWIIAAPPETRIQIQFEDRDP 2268
 QY 1772 FSTEPPN--HDYIETRNGPYETSRMGFSGSELPSLSLSTSHETTVFHSDBSONPGRK 1829
 Db 2269 IEVTPNCTSNVLELRDGDVDDAPILSKFCSTLPSQMSWSGEWMYLRFSNDSPTHVGKR 2328
 QY 1830 LEVQAVYLEQCPREPERANIVKAGY-----NNGQSVTEPCLPGYOLT----- 1873
 Db 2329 AK---SIAQCGGRVPGQSGVSESIGHPI.PYRDN.FCEWHLQGLSGHYLITISFEDFNQ 2385
 QY 1874 ---GHVLTQCOHGTNRMD-----HPLP----- 1893
 Db 2386 NSSG-----CEKDFVEIMDNHTSGNLTGRYCGNTIIPDSIDTSNTAVVRFVTDGSVTASG 2440
 QY 1894 -----KCEVPCCGNITSSNGVYSPQFPSPYSSODCVWLITVPIGHVRL--NLSLQ 1945
 Db 2441 FLRFESSMEBCGGDLOGSIGTFTSPNYPNPHGRICEWRITAPBGRITLIMFNRLRA 2500
 QY 1946 TEBS---GDFTIWDGQOQTAPRLGVTRGM-AKATQSSNOVLKFKHDAAT--GGIPA 2000
 Db 2501 THSSCNNEHYIYANGIRSNSPLEKLCSSVNSNETKSSGNMTKVI.FFTDGSRYG--FT 2559
 QY 2001 IASAVPLTKCPPTLLPNAEVVTENEENFGIDIVARCLPGFTLVGN----- 2048
 Db 2560 ASYTSSEDAVCGGS--LPN---TPRCNFT-----SPGVDGRVNSRLNCEWILS 2604
 QY 2049 -----EITCKLGYLOFEGP-----PRICE 2069
 Db 2605 NPNQGNSSISIHEDFYLESODCOFVLEFRVG--DADGFLMRLCGPSKPTPLVLP 2661
 QY 2070 ---VHCPTNELL-----TD-----STGVLSQSYGSGYPOFTGWMYRV 2106
 Db 2662 YQGWVHFTNRVERHEIGHAKYSFTDCGCIQIGDSGVITSPNYPNADSLTHCWMLEA 2721
 QY 2107 EBDYNISLTVVEYFLSEKO---YDEFEIPDGSQGSPLIKALSGVNSAPLIYTTSSNSGV 2162

Db 2722 POGHTITLTFSPDEIPHTTCAMDSTVTVNNGSPSPSPITIGQICGN-SNPRITQSGSNQLV 2780
 QY 2163 LRNSDHAVNRRKGFKIRYSAFYCSLPRAPLHGFILOQTSTOPGGSIHFCNAGYRLVGS 2222
 Db 2781 VTFNSDHSLOGGF-----YATWNTQT-----LCCGG---IFHS 2811
 QY 2223 MAICTHHPQGHYHMSAIPLOALS--CGLPEAPKN-GMVFGLEYVGTAVYSCSEGYH 2280
 Db 2812 DNGTISPH---WPQNFENSRCSWTALTTHSKLHEISFDNNFLLPSGDGOCOSFVAV 2867
 QY 2281 OGAEBEATBEC-LDTGLMSNRNVPPOCVPTCPDVSISVEHGRWRLIFETQYOFQAOQLM 2339
 Db 2868 WAGTEEVDRKALLATGC--GNVAPG--PYITSNFTAV-----FOSQEA- 2907
 QY 2340 ICDPGYTYTGQVIRCOANGKWSLG-----DSTPTCRIISCBELPIPNGRHIG 2388
 Db 2908 ---PAQGFASFSVSCGSNFTGSPGYIISPNYKQYDNNMNTYV-----LEAN----- 2953
 QY 2389 TLSVYGATAI-PCNSGTVLVSRRVRECMANGLMSGYRCLAGH-----CG--T 2435
 Db 2954 PLSTVLLTFSHLEARSATGS---LVNDS-----VHIIRGYSVMSTPRATVCGDEM 3003
 QY 2436 PEPI-VNGHINGENSYRSGSVVYQCNAGFRLLGMSVRICQODHMSGKTPFCVPITCGHP 2494
 Db 3004 PAPLTIAGVLLNFYSNEQITDF---GFK--FSYR-----ISCQGV 3040
 QY 2495 GNPVNGLOQGNFNLNDVVKFVCNPGY-MAGSARSOCLASGQSDMLPTCRINICTDPG 2553
 Db 3041 FNFSSGI-----ITSPAYSAADYPNDMHLCTYITVSD----- 3072
 QY 2554 HOENSVROVHAGSPHRSFGTIVSYRCHNGFYLTGTPVLSQGDGTWDRPRPQCLVSCG 2613
 Db 3073 ---DKVILKRS---DFDVVPETSCSHDIALI-----YDGANTSD---PLIKFRCG 3114
 QY 2614 HEGSPPHSQMG-----DSTVGAUVRYSCIGKKTIVGNSTRMGDLGHWGTSJPH 2664
 Db 3115 SK-REPNVKSNNMSMLVFKTIDSFOJAKGMKSF--RQTL--GPOQCG--GYLTG--- 3163
 QY 2665 CSGTVGVCGDPGIPAHGIRLRGDSFPDPG-----TYWRFSCEAGHYLR- 2706
 Db 3164 ---SNMTPASBPDSNSM-----YDKNLNCWIIIAIPVNKVIHLTFNTFALBASTRQR 3214
 QY 2707 -----GSSERTCOANGSMGSOPECGVISCGNPGTSPNARVVSDDLAVSSSIVY 2756
 Db 3215 CLYDVYKLYDDSDENANLA-CTPCGSTVPAPFISGNFLT-----VQIISDLTL 3262
 QY 2757 ECRREGYAT-GLSRHC--SVNGTWG---SDEPCLVINCGDPGPIPANGRLGNDFRYNK 2810
 Db 3263 E-REGFNATYTIMMPCGGTYNATWTPONISSP-----NSSDPDVPFS- 3304
 QY 2811 TVTYQVPGVYMMESHRSVLSCTKDRTMNGTKPVCKALMCKPPLINGKVVSGSDPMWGS 2870
 Db 3305 -----ICTWVDSPPH--OOVKIT-----VWAL 3325
 QY 2871 SVT-YACLEGYQLSLPAVFTCEGNGS---WTGELPOCFPVFCGD-----PGVPSR 2916
 Db 3326 QLTSDQCTQNY---LOQDSBQGHGNGRFOFCGNNAVAVVFVSSMSTAMWIFKSGVNR 3382
 QY 2917 GRREDRGPYSR--SSVSISCHPPLVTVSPPRFQCSDBTSGTQSCIDPLTLTTCADGPV 2975
 Db 3383 NSR--MSFTYQIADCNNDYHKAFCNLRSP-----GMPDNYDNDXDCVTJTLA---P 3428
 QY 2976 Q-----FGIUNNSOGYQVGSIVLFRCKQGYL-----LOGSTRTCLPN- 3013
 Db 3429 QNHTISLFHSLIGIENSVE-----CRNDFLEVNRGNSNSPFLKCYGCTLLPNP 3477
 QY 3014 ---LWSTGTPDPCVPHHCQBPPTPHANYGALDLPSMGY 3049
 Db 3478 VFSQNNELYLRFKSDSVTSDRGEYIIMTSSPSGC-----GGLYGDREGSFTSP--GY 3527
 QY 3050 ---TLTPARR-----ASPS----- 3061
 Db 3528 PGTYPNNTYCEWVLVAPAGHLVITINFFYFISIDDPGOCVQNYTLTYDGPANASSPSSGPGYCG 3587

Oy 3062 ---RVAPSTADAP-----RMAAGQASRPSAMR 3085
 Db 3588 GDTSIAPFVASSNOVIFKFMADYAKRPSAFR 3618

RESULT 11

OyTUS3 PRELIMINARY; PRT: 3620 AA.

AC 09TUS3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Cullin.
 GN Name-CUBN;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_Taxid=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=20021710; PubMed=10552972;
 RA Xu D., Koziyaki R., Newman T.C., Fyfe J.C.;
 RT "Genetic evidence of an accessory activity required specifically for
 RT cullin brush-border expression and intrinsic factor-cobalamin
 RT absorption.";
 RL Blood 94:3604-3606(1999).
 DR EMBL; AF137068; AAF14258.1; -.
 DR HSSP; P35555; IEMN.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR InterPro; IPR000152; Aex_hydroxyl_S.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_like.
 DR Pfam; PF00431; CUB; 27.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF07645; EGF_CA; 3.
 DR SMART; SM00042; CUB; 27.
 DR SMART; SM00179; EGF_CA; 4.
 DR PROSITE; PS00010; ASX_HYDROXYL; 3.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 5.
 DR PROSITE; PS01187; EGF_CA; 4.
 DR EGF-like domain.
 SO SEQUENCE 3620 AA; 397432 MW; 00B041EB6AD07348 CRC64;

Query Match 8.6%; Score 1457.5; DB 2; Length 3620;
 Best Local Similarity 21.4%; Pred. No. 6.5e-74;
 Matches 679; Conservative 344; Mismatches 1070; Indels 1073; Gaps 117;

Oy 113 TCMKYSDEFAA-----WSDHRVPCARWCDALHGPSGIITSPNPILOYDNNACHWIT 167
 Db 443 TCVGVANAFSCCECTRFMTGFLCQIPQVCGSLSDMSGSFYSMDGVYVDVNCFWIR 502
 Oy 168 ALNSKVIKLAFEEDFLER---GYDTLVGDGDGDQDKTVLYWSQACSDSPHTPSR 223
 Db 503 T-EDRKVLRIFTTFQLESVNNCPHEFLQI---HDGSSAALQIGRCGSLVLRH----- 552
 Oy 224 IPESMSGDIMRQKKTVLEICRDISSDARSVKSPKTSNAVELVAPGTEIEQSGCDP 283
 Db 553 --ELIS-----SNNAL----- 561
 Oy 284 GIPAYGREGSRPHNGDTLKECPAPAFELVGKAITCQKNNWSAKKPGCVFSCFENFS 343
 Db 562 -----YFLVSEHRSRGRTI-----RMETQOECGGLMGTY-- 595
 Oy 344 PSGVLLSPNYPEDYGNHLICWMLILARPESRIHLAFNDIVER---QFDFLVIKDGATA 399
 Db 596 --GSIKSPGPGVNPVRGRDVCWQVVTSPDLLITFTFGLSLSHHDDCKDVLIRDPFLY 653

Oy 400 EAPVLGFSGNQLPSSITSSGHVARLEFQNDHSGKRGFNITFTT----- 444
 Db 654 QDPSELGKCTTLLSVPLQTTGPPFAKRVHSHSNQINDGGFHITVLTSPSDLHCAGNYDPE 713
 Oy 445 -----FRHN-ECPPDGVPNVGRFGDSLQLG-----SSISFLCDEG 479
 Db 714 GLSSDLSGPFTHNRQC-----IYIIKQPLBEGIQVNTVHLEBQSSCSQSHIEVRDKI 769
 Oy 480 FLG-TQSEITTCVLEKGSVW-----NSAVLRC-----EAPCGHILTSPSGTILSPGW 527
 Db 770 LKGVACNETLPHIKSIRNHIMIKIDASILVASFPAVYQVACGELTG-EGVIRSPFY 828
 Oy 528 PGFYKDALSCAMVLEADPGYRIKTFPRFKTE-----VNYDLEVRDGRITSAPILGYVHG 583
 Db 829 PNVPYGERICKWTIHQPOSOVYINLFPAGFESSAHCDTDVIEIGSSSILGSPENKRYCG 888
 Oy 584 TQVQPLISTSNLYLLFSTDKSHSDIGFOLRYETITLQSDHCLDPCI PVNGGRHNDIFY 643
 Db 889 TDIPLFITSVNPLVYIVFKSSSTENHGFMAKFS----- 922
 Oy 644 VQALVTSCDSGYTLSDGEPLCEPWFQWSRALPSCALCGFTIGSSGTTILSPGPDY 703
 Db 923 -----AADLACGEILTESTGIIQSPGHPIY 948
 Oy 704 PNNLCWTIITSHGKGVFTFHTFHE-----SGHDVLLITENSGFTQPLRQLTGSRLPA 759
 Db 949 PHGINCTWHILVQGHILHLIFRFHLEFHYNCTNDLEVYDTSNTY-LGRYCKSIP 1007
 Oy 760 PISAGLGNFPAQVRFISDFSMSEYEGNITFSEBDLPCEBEPVAYSIRKGLQFGVODT 819
 Db 1008 SLTS---STNSKILFIADSDLAIEGFLINYEADAS----- 1041
 Oy 820 LTFSCFPGYRLGEGTARITCLGRRRLWSSPLPRCVACGNSVTGQTTLSNPNVYNN 879
 Db 1042 -----SACMEDYENSGTFTSPNPNVYNN 1066
 Oy 880 NHECIYSIQTPGKGIQAKARFELSF-----GDVLKYDDGNMNAARLLGVSHSEBM 932
 Db 1067 NMKCIYRIVETSOQIALHFTNFALBEAIGQCVADVEIIRDDGETSPPLGYCGS-IP 1125
 Oy 933 GVTNLSSTSSIMLDFITDAENTSKGFLHFSFELIKCEDPGTPEKFGYVHDEGHPAGSS 992
 Db 1126 PPIIHSNKLMLQFTSD-----FLGSG 1148
 Oy 993 VSFSCDPPGYSLRGSEBELLCSGERRVDRPLPTCVACGGTVRGVSGQVLSPGYPAYE 1052
 Db 1149 -----PGFS-----AYWDGSL---TCCGGNIT-TPTGVFTSPSYPMPLY 1183
 Oy 1053 HNLNCITTEAEAGCTTGLHFLVPTDE---VHDVLRINDGPVSGVLLKELSGPALPK 1108
 Db 1184 HSSECYWLLKXSHSPPELEBEDHLEHNPCTLDYLAUYDGPSTSHLLSOLCGNKRPP 1243
 Oy 1109 DLHSTFNSVYLQSFDTDFTSKQGFALQF-----SVSATSCNDPGIQQNGSR 1156
 Db 1244 VIRSTGSMFLKFTTDDQCGGFLAKYQOTCRANVIVNKNYGLISEIHINPNPSDQRC 1303
 Oy 1157 GDSWEA-GDSTVFOCDRGVALQGSARISCVKIE-----NREFWQPSPT----- 1199
 Db 1304 NMTIQAATGNTVNTVFLAFELHNHINCSYDLELYDPRRMGRYCGADMPTGTTGSKL 1363
 Oy 1200 -----CIAFCGDLTGPSGVILSNPNYPRPYPRGKCDMKVYVYSPD 1239
 Db 1364 QVLFYTDGVGHQEGFQWQFHHQCGGELGTGTSFSPGVPNTYPPNKECTIWTITAPG 1423
 Oy 1240 VYIALVFNIFNLB---PGYFLIHYDGRSLSPUG---SPYSGQLPGRILSSNSLFL 1292
 Db 1424 SSIGLTHDHPVEVHAACNFDVLEVYGGPRPHSRITIQLSQSSSNPMQVSSIGNELAI 1483
 Oy 1293 AFRSDASVSNAGFVIDYTENPRESCFDPSIKNGTRVGSCLKLGSSVTYVYCHGVEVGT 1352
 Db 1484 RFTKDSINSINRGFVNASQAVP-----GG----- 1506

QY 1353 STLSCILGPDGKPVNNRPVCTABCGGQYVSDGVLSPNVQNYTSGQICLYFVTVPK 1412
 Db 1507 -----CGGIFQAPNKEIHSPPYPSYKRENTCCSWIRVER 1541
 QY 1413 DYVEGQAFRITLAND-VVEYHGHSHSLSSLSGHTGESLPLATSNQVLIKESAK 1471
 Db 1542 NHRILANTDPELDQSDCITAYDGLSSTTRLASVCGRQOLTNPITSSGNSLFLRQSG 1601
 QY 1472 GLAPRGFHYQAVPRTSATQCSVPEPRYGKRLGSDFSGALVREPCNSGYALQSSPE 1531
 Db 1602 PSRQGRGFR----- 1610
 QY 1532 IECLVPFGALQWNSAPTCVPCGN-LTERRGITLSPGPPEBYLSLNCWMI-VVPE 1589
 Db 1611 -----AQFN-----QVCGGHILTNSPDTISPLFPAYKPNQNSWITQAP 1653
 QY 1590 GAGIQIYVSVFTQN-----WDSLSEVPFGADNTYTMIGSFSGTVPALLNSTSNQVLYHF 1645
 Db 1654 FNHITLSPDHFGLESSTCTQDPLEILDGDYDAPLGRYCGHSMHPHITSPSGLTLRF 1713
 QY 1646 YSDISVSNAGFHLKYKTVGLSSCEPAVPSNGVTKGRYLNDVVSFOCEPGVALQSHAH 1705
 Db 1714 VSDSRVNSDGFHATY-----AASSGA----- 1734
 QY 1706 ISCPMGTVRRMNYPPPLCIAQCGGTVEEMEGVILSPGPNGYPSNDCSWKIALPVGFGA 1765
 Db 1735 -----CGGTFFMAEGIFNSPGYPEYVPSNVCWNIYSSPGRNL 1773
 QY 1766 HIOFLNFSTP-----NHDXIEIRNGPYETSRMRKRESGSLPSLS-TSHETTVYFHS 1820
 Db 1774 QLSITITQLEDSQCSDFVEVRBG-NATGHLVGRYCGNVLPLNYSIVGHILMIRVSD 1832
 QY 1821 HSQNRPGFK----- 1830
 Db 1833 GSGGTGFQATFTKIFGNDNIYVTHGKIAPMPGRYHNSQWIMVNAVATQVHGRIL 1892
 QY 1831 EYQAYEIOEC-PDEPRFANGI--VKGAGYVNG-----QSVTECLPEYQULTGHP 1876
 Db 1893 EIDIEGASQCYDRLARYDGLGHSRLIGYCGTQTSPSSSNNSLTFQPSDSITGKG 1952
 QY 1877 VLTQHGNTNRMDHPLPKCEV-PCGANTSSNGTV--YSPGFSPSSSQDDCWLTVPI 1933
 Db 1953 FILEWFAVMS-GGRLPTATGACGFLRTGDAVPLFSPGWPEBSYNSNDCWMLQAP- 2010
 QY 1934 GHGVRNLNLLQTEPSG---DETTIWDGPOQAPRIGVFTSRMAKTVQSSSNQVLYLF 1989
 Db 2011 DSTVELNIILSDIEAGRTCDYDKLVRDGDNLAPQAVLQGREIPIRSTGEWFIHF 2070
 QY 1990 HRD-AATGIFALAF-----SAVPLTKCP--PPTILPNAE-----VVT 2024
 Db 2071 TSPFSTTAGAFNASFHKSCGYLHADRGITTSPOYEPETYSPLNCSMHVAVOSGLTIAVH 2130
 QY 2025 ENEEFNI--GDIARYRCLPGFTLV--GNEILTCCKLTQIO--FEGPPICEVHOPNTE 2076
 Db 2131 FEGPFOIPSD---SSCSQSDYLVLNKGNDPIYSPRIGRGNHFGCSPPSSTLFTSDQ 2187
 QY 2077 LL-----TDSITGVILLQSPSYPOFQTCWTL 2103
 Db 2188 MFVQFISDGSNGQGFKEYEAKSLACGNIVYHDVNSAGYVTSPPHNNYVQHDQWML 2247
 QY 2104 VRVEPDYNIISLVEYFLEKQYDEF-----EIPFGPSQSLALALSGNTSA 2150
 Db 2248 IAPPFKILRQVE-----DOFNIETPNCVSNYTLRQGVDSNAPLAKLQGR-SL 2298
 QY 2151 PLIVTSSNSVYLWMSDHAYNRKGFIRYSAFYCSLPRAPLHGFLIGTSTQPGSGIHF 2210
 Db 2299 PSSQSSSGVWYLRFRSDNSSTQVGFKIKYATAC----- 2333
 QY 2211 GCNAGRLVGHSMALITRHPQGHMWSBALPQLQALSCGLPAPAKNGVFGKEYTVGTKA 2270
 Db 2334 -----GGKRVYGQSGII-----ESSGYPTLPYRDNSTFCGMHLKGPSSG 2369
 QY 2271 VYSC--SEGVIHQAGAEATABACLDGTGLMSNR--NVPPQCVVTCPD----- 2312

Db 2370 HYLTIFEDHLQNSGCKEDVE--IWENHTSGNLGKYCGNTTIPDSIDTSSNALVRF 2427
 QY 2313 VSSISVEHGRWRLIFETQY-----QFOQLMLICEPGYYTQGVIRCOANGKMSLGDST 2367
 Db 2428 VTDGSLTAGAGFLRFESSMEACGGELQGTGTFTSNY-----PNRN 2469
 QY 2368 PTCRIISCEBLPPPNHIG-TLSVYGATAIPSGNGYTLVGSRYRECMANGLSGSEV 2426
 Db 2470 PHGRV--CEWRIMVOGRRITLTFFNNLRLAHPSCVSEBVTI-----FNGIRNNSPO 2519
 QY 2427 RCLAGCGTPEPIVNGHNGENYSYRGSVYQCNAGFRILGMSVRIQQDDHMSGTPPC 2486
 Db 2520 --LEKICGS-----VNASSEIKSSGNTMKVFTDSDRPPGFSAT-----YTSSEDAVC 2567
 QY 2487 VPIITCGHPNPVNGLTQGNQFNLNDVVKVPCNPGYMAEGAARSOCLASQWMDMLPTCRI 2546
 Db 2568 -----CGSLJHPPEGN-----FTSPGY--NGVSNYS-----RN 2593
 QY 2547 INC-----TDPGHOENSVRQVHASGPHRPSFGTTVYRCNHGTYLLGTPVLSQGD----- 2597
 Db 2594 LNCWTLNRP--NOGNSIYIHF-----EDFYLSHQ--DCQFDVLEFR 2633
 QY 2598 -GTWDRPRPQCLVSCGHPSPPHSOMSGDSYTVGAVVRYSCIGKRTLVGNSTRMCGIDG 2656
 Db 2634 VGNADGFLMRL--CG-PSKP-----IYPLVPIYEPWIFHFTNEHEHYGF-- 2677
 QY 2657 HMTGSLPHSCGSTV--GYCCGPGIPIAHGIRLQGSFDPGYVWRFSCEA--GHVLRGS-- 2708
 Db 2678 HAEVFTDCGGIQLDSGVIASPNYPA-----SYDLTCHCSWLEAPQGFITILTSD 2730
 QY 2709 -----SERTCQANGSSSQPREGVISCNPGTP-----SNARVVSQD--LVSS 2752
 Db 2731 FIDEDHATC--ANDSVVRNG-----GSPSPITIGYCGTSPPTIQSGSNQVLVIFNS 2782
 QY 2753 SIVECREGYAT-----GLISRHCSYNGT-----WTGSDPECLVINGDPGIPAN 2798
 Db 2783 DHSVQ--NGGFAYIWMQTQGLGCGGILH--SUNGTRIRSHMQNPE--NSKCSWTVITHE 2836
 QY 2799 GLRLANDRNYKTVTY--QVPPGYMESHRVSVLSTCKDRTMNGTKVPCALM--CKP 2852
 Db 2837 SKQLEISFDNNFRIPSGDQCONSIV-----KVMAGTEBVAESLSTATGC-- 2880
 QY 2853 PPLIPNGKVGSDPFMWSSTVYACLEGYQ-----LSLPAYVTCGN 2893
 Db 2881 -----GNVAPGSILTPRNVIAVFOCETPAQGSASFVSRCCGN 2920
 RESULT 12
 ID 0723G3
 AC 0723G3 PRELIMINARY; PRT; 1441 AA.
 DT 01-OCT-2003 (TREMblrel. 25, Created)
 DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Hypothetical protein DKFZp686i10142 (Fragment).
 GN Name=DKFZp686i10142;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Human colon endothel primary cell culture;
 RA Koehler K., Beyer A., Mewes H.W., Weil B., Amdt C., Osanger A.,
 RA Poto G., Han M., Wiemann S.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 CC -! SIMILARITY: Contains 3 EGF-like domains.
 DR EMBL; BX537918; CAD97901.1; -
 DR InterPro; IPR000742; EGF_2;
 DR InterPro; IPR006209; EGF_like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00008; EGF; 3.
 DR Pfam; PF00084; Subf; 22.
 DR SMART; SMART; 22.
 DR SMART; SMART; 22.
 DR SMART; SMART; 22.
 DR PROSITE; PS00022; EGF; 3.
 DR PROSITE; PS01186; EGF; 2.
 DR PROSITE; PS00026; EGF; 3.
 DR PROSITE; PS00023; EGF; 2.
 DR EGF-like domain; Hypothetical protein.
 FT NONTER 1
 SQ SEQUENCE 1441 AA; 157497 MW; BA8B04EC59962A2A2 CRC64;
 Query March 8.3%; Score 1416; DB 2; Length 1441;
 Best Local Similarity 25.1%; Pred. No. 4.4e-72;
 Matches 407; Conservative 179; Mismatches 591; Indels 444; Gaps 52;
 QY 1668 CPEPAVSNVKTGERLYLVNDVVSFOCEPFGALQGHANISCPGTVRMNPRLCTAOC 1727
 DB CGEPSPIMNGVAGSNYSFGAMVAVSCKNGFYIEGKSTC--BATQWMSPIPTC---- 69
 QY 1728 GGTVEEMEGVLSFGPGNTPSNMDCSKYKALPVFGAHIOPLNFSTEPNHDYIEIRNGP 1787
 DB 70 -----HPVSCG-----EPP-----KVENGF 84
 QY 1788 YETSRMMGRFSGSELPSLSLSTSHETTYPHSD--HSQNRPGFK--LEYQAVELQ----- 1838
 DB 85 LE-----HTTGRIESEVRVYQCNPGYKSGSFVFCQANRMHS 123
 QY 1839 -----ECRDEPANGIVRGAGVNGSVTFECLPGYOLGHPVLTCQHTNEMD 1889
 DB 124 ESLPMCVPLDCGKRPRIQNGFMKGBNFVSGKVOFCNEGYELVGDSSMTQ--KSKGMN 181
 QY 1890 HPL-PKC-----EVPCGN-----ITSSNGTVYSGFSPYSSQDCWLTIVPIGHG 1936
 DB 182 KKSNNPKMPKACRPPLLENQLVLKELTEVGVV-----TFSCKE-----GH- 223
 QY 1937 VRLNLISLQTEPSQDFTITWDGPOQTAARLVGTFRMAKKTIVGSSNQVLKPHRDATG 1996
 DB 224 VLQGPVVKCLPSQO-----NM----- 240
 QY 1997 GIFAIAFSAYPLTK---CPPTILPNAEVTENEENIGDIVRYCLPGFTLVGNEILTC 2053
 DB 241 -----DSFVVCIVCTPPRL--SGVPIPSALHFGSVKXSCVGFPLRNSITLIC 292
 QY 2054 KL-CTYIQFBGPPIC-EVHCPTNBLTDSGTGVLSSQSPESYQFOTCSMLVAVEPDYN 2111
 DB 293 QPQGTW---SSPLPECVPECPOPEEIPN--GIIDVQGLAYLSTALYTC-----KQGF 341
 QY 2112 ISLVEYFLSEKQYDFEIPDPSGQSPFLKALSGNYSAPLIVTSSNSVYLKSSSHAY 2171
 DB 342 L-----VGNITTLGCGENGW----- 356
 QY 2172 NRKGFIRYSAPYCSLPAPLHGFILGQSTQPSGSIHFGNAGYRLVGHSMALCTBRPQ 2231
 DB 357 --LGKPTCKAIECLKREILNG--KFSYTDLHYQIYVYSCNRRFLREGPALLCTLETD 413
 QY 2232 GYHLSSEALPLCOLSGLEAPAPKNGVFGKEYVGTGKAVYSGEGYHLDGAETACBL 2291
 DB 414 ---WDVAPSCNAIHCDSPQIENGFEAGADYSGAIIIVSCPPG--QVAGHAMQTC 467
 QY 2292 DTGLMSRNVPQCVPTC-----PDVSSISVENG----- 2321
 DB 468 ESG-WSSS--IPTCMPIDCGLRPHIDREDCTKLXDDQYFEQEDDMEEVRYVTPHRYHL 524
 QY 2322 -----RRRLTFE-----TOYFOALMLICDPGYUYTGQVIRICQANGKSLDSTPTC 2370
 DB 525 GAVAKTMENTKESATHSNLYGTWVSYTGNPGYELLGNVLLCQEDGTN--GSAFSC 582
 QY 2371 RIISCGELIPPNHRIIGTSLVYATAI FSCNSGYTLVGSFVRRCMANGMLSGSEVRCLA 2430
 DB 583 ISICCDPLTAPENFLLFTETSMGSAVOYCKPGHIIAGSLRLCLERKMKSGAPRCEA 642
 QY 2431 GHCTPRPIVNGHNGENYVRSVGVVQCNAGPRLIGMSVRLICQDDHMSKTFPCVPIIT 2490

DB 643 ISCKKPPVANGSISKSNTYTLSTLYECDPGVLTNTERTCQDDKNWDEDEICLPVD 702
 QY 2491 CGHPENVNGLTQGNQENLNDVVKFVCPNGYMAAGARSOCCLASGQSDMLPTCRIINCT 2550
 DB 703 CSSPPVANGQVGDDEYTPQKEIEHTCNBEGFLLEGARSVCLANGSWSGATPPCVPVCA 762
 QY 2551 DPHQENSVPQVASHHRSFGTYSYRCNHPFYLLGTVLSCQGGTMDRBRPOCLV 2610
 DB 763 TPQLANGVTE---GLDGFMEKVEYFHCHGILLHGAFLKLTQSDGNWDAEIPLCXPV 817
 QY 2611 SCGHPGSPHSGMSGDSYTVGAVVRSYSCIKRTLVGNSTBMC----- 2652
 DB 818 NCSEPDLANGFNGFNGFNGHGHQYOCFPGYTLHGHSRCLNSNGSSSSPCLPCCR 877
 QY 2653 -----GL-----DGHWTGSLPHCSGTYSVC 2673
 DB 878 STPIEYGTVNGTDFDCKAARIQCFKGLILGISBITCBADQMSGSPHCEHTS---C 934
 QY 2674 GD-PGIPAHGIRLGDSPDPTWRPSCBAGHYLRGSEERTQANGSSSQPBGVATSCG 2732
 DB 935 GSLPMIPNADISSTSWKE--NVITYSCRSQYVIQSSDLICTEKGVMSQPYPCBPLSCG 993
 QY 2733 NPGTPSNA----- 2740
 DB 994 SPSPVANAVALTGEAHYTESVYKRLCLEGYTMDITDITTCQKQGRMPERRISCPKRCPL 1053
 QY 2741 -----RVFSDGLVFESSIVYECREGYATGLSRHCSVNGTWG--SDPECLVINGCD 2792
 DB 1054 PENITHTLVHGDPSVNRQVSACBEGYFEGVNI SVQDGTWEPFSDSCSPVSCGK 1113
 QY 2793 PGIPANGLRGNDPRNKYTYTQCVBYMMESHRVSLSTKORTNNGTAPVYCALMKCP 2852
 DB 1114 PESPEHFVVGSKYTFESTIITYOCEPEYLEGNRERY--QENRQMSGVAICETRCET 1171
 QY 2853 PPLIPNGKVVGSDPFMMSSYTYACTLEGYQLSLPAFTCEGNSWTGELPOCFPVFCGDPG 2912
 DB 1172 PLEFLNGKADIERRTGTPNNVYSCNKGYSLEGSSEAHCTNGTSHVPLCKRPPCVPF 1231
 QY 2913 V-PSRGREDRGFSYSSVSFSCHPPLVLVGSPPRFQSDGTWSTQPSCIDPTLTTCAD 2971
 DB 1232 VIPENALLSEKEFVDQNVISIKREBFLQGHGIIITCNPDETWOTSAKC---EKISGCP 1288
 QY 2972 PGVPQEGIQNNSQGYQVGSIVLPRCQGYLLQSTTRTCLPNIWSTSPDPVPHHROP 3011
 DB 1289 PAHYENAIARGVH-YOYGDMITYSQVYMLBEFLRSVCLENGWT--SPPIC-RAYGRFP 1345
 QY 3032 ETPHANVALDPS-----MGYTLIPARRASPRVAPSTAPARMAAGQASRPSA 3083
 DB 1346 ---CONGGICQRPNACSCPEGMGRLC-----EPPICILPCLNGRCVAPYQCCDPPG 1395
 QY 3084 W 3084
 DB 1396 W 1396
 RESULT 13
 Q8NB79 PRELIMINARY; PRT; 1497 AA.
 AC Q8NB79;
 DT 01-OCT-2002 (Tremblere1.22, Created)
 DT 01-OCT-2002 (Tremblere1.22, Last sequence update)
 DT 01-MAR-2004 (Tremblere1.26, Last annotation update)
 DE Hypothetical protein FLJ90754.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Iisagai T., Ota T., Nishikawa T., Hayaashi K., Otsuki T., Sugiyama T.,
 Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,

Yamamoto T., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 Maehara Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 Hatohori A., Okumura K., Iwayanagi T., Ninomiya K.,
 CC Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RL -1- SIMILARITY: Contains 3 EGF-like domains.
 DR EMBL: AK075235; BAC11489.1; -.
 DR HSRP: P00740; IEDM.
 DR GO: GO:0004872; F:receptor activity; IEA.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR006209; EGF_1k.
 DR InterPro: IPR006210; IEGF.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00009; EGF_3.
 DR Pfam: PF00084; Sushi; 23.
 DR SMART: SM00032; CCP; 23.
 DR SMART: SM00181; EGF; 3.
 DR PROSITE: PS00022; EGF_1; 3.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS00026; EGF_3; 2.
 DR PROSITE: PS09923; SUSH1; 23.
 DR EGF-like domain; Receptor.
 KM EGF-like domain; Receptor.
 SQ SEQUENCE 1497 AA; 163802 MW; 35B645633C733B CRC64;

Query Match 8.3%; Score 1415; DB 2; Length 1497;
 Best Local Similarity 24.9%; Pred. No. 5.2e-72;

Matches 405; Conservative 183; Mismatches 615; Indels 426; Gaps 49;

QY 1486 VERTSATQSSVPPRYG---KRLGSDSVGAIVFECNSGYALQSGPEIECTLPVGCALA 1542
 DB 1 MPRCIAHFCEKPEFSVSYLESVAKFAAGSVVFCMEGFLVNTSAKIECH---RGC 56
 QY 1543 QMNVASATCVPCCGNLTERGRTLLSPGPEPPIVNSLNCWKIVVEGAGIQIIVSVPT 1602
 DB 57 QMN-----PSPM---SIQC----- 67
 QY 1603 EQMNDSEVDFGADNTVTMLGSEGTVPALLNSTNQYLHFYSDISVAGFHELYKT 1662
 DB 68 -----IPV----- 70
 QY 1663 VGLSSCEPVPNSGVKTERYLVDVVSFOCEPFGVALQGHAIISCMPTGRMKNYPPPL 1722
 DB 71 -----RCCEPSPIMNGVAGSGNYSFGAMVAVSCNKGFIYKGEKSTCEAS--GQMSPIPT 124
 QY 1723 C-IAQGS-----GTYEMEGVIL-----SPGEPGVPYNNMCCSKIALPVFGAH 1766
 DB 125 CHVSGCEPKNKENGFLHTTGRIFSEVARYQCNPGYK-----PVSGPVF 169
 QY 1767 IQPLNSTEENHDYIEIRNGPYETSRMGRFSGSELPSILSTSHETTYVPHSDHQNRP 1826
 DB 170 VCHAN-----RHMSESPLMWVPL----- 188
 QY 1827 GFLLEYQAYELQCEPPEPPANGIVRGAGVNGQSTFECLEPGYQLTGHPIVLTQHNTNR 1886
 DB 189 -----DCGKRPPIQNGFMKGEMFEVSKQVFCNMGSEYELVAGSSTWCQ--KSG 234
 QY 1887 NMDHPL-PKC-----EVPQGN-----ITSNGTVYSGPSPVSSQDCVMLITVPI 1933
 DB 235 KMKKSNPKCMKAPKCEPPLBNQYLKELTEVGV-----TFSCGE----- 277
 QY 1934 GHGVRLNLSLLQTEPSGDFITTDGPOQTAIRLGVFTSMARKTVQSSNQVLKFXHDA 1993
 DB 278 GH-VLQGPSTVLKCLPSEQ--WN----- 296
 QY 1994 ATGGGIAIAAFSAVPLRK---CEPPTILPAAEVVTENEFNIGDIYRIKRLPGFTLVGNEI 2050
 DB 297 -----DSFPVCKIVLCTPPLLI-SFGVPIBSALHFGSTVYKSCVGGFFLRGNS 345
 QY 2051 LTCKL-GTYLOFEGEPPI-C-EVHCTNELITDSTGYISQSVGSGYPOFGTCSMLVRYEP 2108
 DB 346 TLQCPDGTW---SSPLPECVIVBCPQPEIIN--GIIDVKGALVSTALYTC-----KP 394
 QY 2109 DYNISLTVYFUSEKQYDFEELFDGSGGSPILKLKSGVSNAPLIVTSSNSVYLKRWSSD 2168

DB 395 GFEL-----VQNTTLLCGENGM----- 412
 QY 2169 HAYNRGFKIRISAPYCSIPRAVPLHGFILGOSTPQSGSIHRCNAGYRLVGHMAICTR 2228
 DB 413 -----LGGKPTCAIECLKPKELING-KFSYIDLHGYQIVTVYSCNNGFLRGPSTLCIE 466
 QY 2229 HPQGYHLMSEALPLCOALSCGLPEAPKNGMVGKEXYTVTKVYVSGSEGYHQAQAEATA 2288
 DB 467 TGD-----WVDABSCNAIHCDSPQPLENFVAGADYSYALITVSCPPG--QVAGHANO 520
 QY 2289 ECLDTGLMNRNVPPQCVPTC-----PVSISYVHG----- 2321
 DB 521 TCEESG-WSSS--IPFCWPIDCGLPPIHDFGCTKLKODQYFEGEDDMVEVYTPHP 577
 QY 2322 -----RWRLIFE-----IQYQQAQMLICDPGYTYTQRYVIRCOANGKMSLGST 2367
 DB 578 YHLGAVAKTWENTKSPATHSSNPLVGTWVSYTCNPGYELLNPNALICQEDG--TRNGSA 635
 QY 2368 PTCRIISCGELPIPNNGHRIGLSYGATAIPSCNSGYTLVGSRRVREGANGLMGSEVR 2427
 DB 636 PSCISIECDLPTAPENGLRFTETSMGSAVQYSCKPGHILVSGRLCLENRKMSGASPR 695
 QY 2428 CLAGHGTPEPIVNGHINGENYSYRGSVVYQCNAGFRLIGMSVRIQQDHHWSGKTPCV 2487
 DB 696 CEAISSCKENPYMNGSISIGSNYTYLSTLYECDPGYVINGTERFRCTODKMWDEDEPICI 755
 QY 2488 PITCGHPGNPVNGLQGNQFNLDVYKFCNPGYVABEAAASOCCLASQMSMLPTCKII 2547
 DB 756 PYDRSSFPVANSAGVQGDYITFOKEIETYNCEGLFLEBARSRVCLANSWSGATDVCV 815
 QY 2548 NCTDGHQENSVYQVHASGPHRFSPGTVSYRCNHFYLLGTPLYSCQDGTWDRPRQC 2607
 DB 816 RCATPQLANGYTE-----GLDYGMKEVTHCHGEGYILHAPKLTLCQSDGNWDEIILC 870
 QY 2608 LTVSGCHGSPHSGMSDSTVGAUVARYSCITGRKTLVGNSTRMGLDGHWTGSLPHCSG 2667
 DB 871 KPVNCGPEPDLAHGFPNGFSFIHGHIQYQCPGPKLHGNSRRLCTSNWSGSGSPCLP 930
 QY 2668 TSVGVCGDPGIPANGIRLGDSEFDPQTVWRFCSEAGHYLRGSSERTCOANGSMGSGQPRCG 2727
 DB 931 FR---CSTPVL-EYGVNGTDPDCKKARIQCFKGLLGSEITTCADGQMSGFPICE 986
 QY 2728 VISGNPGETPSNARVPSDGLVFSSSIY-YRCREGYATGLLSRHSVNGTWTGSDPECL 2786
 DB 987 HTSCGS--LPMIPNAFISSETSMKENVITYSCRGYVYQSSDILCTEKGWSPQPYRCE 1044
 QY 2787 VINCGDPGIPANGIRLGDSEFDPQTVWRFCSEAGHYLRGSSERTCOANGSMGSGQPRCG 2846
 DB 1045 PLSGSPSVANAVATGEAHTYSEVVKURCLEGTYTMDT-TDPTCQKDGMRPBRISCS 1103
 QY 2847 ALMKCPPLIPNGKVVVSDPFMKGSSVYACLEGYQLSLPAVFTCEGNGSWTGLP---Q 2902
 DB 1104 PKKCPPLPENIHLVHGDGFSVNRQVSVSCAEGTTFEGVAINSVQLDGTW--EPPFDES 1161
 QY 2903 CFPVFCGDPGVPSGKREDKFSYRSSVSCHPPLVLSPPRRFCOSDGTWSTGPSCT 2962
 DB 1162 CSPVSCGPSPESPEHGFVGSKYTFESTIIVQCEGYELRERARVCOENRQMSGVAIAC 1221
 QY 2963 DPTLTTCADRCVPOFGIIONNSQGVQVSTVLFRCQKGYLLQGSTTRTCLNLTMSGTPPD 3022
 DB 1222 ETRCETOLEPLNGADIENTNT---TQPNVVISCNKSYSLGSEBAHCTENGTSHPVL 1277
 QY 3023 CVPHHCRP 3031
 DB 1278 CKPNPCVP 1286

RESULT 14
 070244 PRELIMINARY; PRT; 3623 AA.
 AC 070244;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-JAN-1999 (TREMblrel. 09, Last sequence update)

DT 01-MAR-2004 (TREMELREL. 26, Last annotation update)
 DE Intrinsic factor-B12 receptor precursor.
 GN Name=CUBILIN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 NX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96148073; PubMed=9478979;
 RA Mostrop S.K., Kozyraki R., Kristiansen M., Kayen J.H.,
 RA Rasmussen H.H., Brault D., Pontillon F., Goda F.O., Christensen E.I.,
 RA Hammond T.G., Verroust P.J.;
 RT "The intrinsic factor-vitamin B12 receptor and target of teratogenic
 RT antibodies is a megalin-binding peripheral membrane protein with
 RT homology to developmental proteins."
 RL J. Biol. Chem. 273:5235-5242(1998).
 DR EMBL; AF022247; AAC71661.1; -.
 DR PIR; T08618; T08618.
 DR HSSP; P00740; 1EDM.
 DR GO; GO:0005599; F:calcium ion binding; IEA.
 DR GO; GO:0004872; F:receptor activity; IEA.
 DR InterPro; IPR000152; AaX_Hydroxyl_S.
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR Pfam; PF00431; CUB; 27.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF07645; EGF_CA; 3.
 DR SMART; SMO0042; CUB; 27.
 DR SMART; SMO0179; EGF_CA; 4.
 DR PROSITE; PS00010; AaX_Hydroxyl; 3.
 DR PROSITE; PS01180; CUB; 27.
 DR PROSITE; PS00022; EGF_1; 4.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS00026; EGF_3; 6.
 DR PROSITE; PS01187; EGF_CA; 4.
 DR EGF-like domain; Receptor; Signal.
 FT SIGNAL 1 20 Potential.
 FT CHAIN 21 3623 Intrinsic factor-B12 receptor.
 SQ SEQUENCE 3623 AA; 398984 MW; 39FB792AC6545240 CRC64;
 Query Match 8.3%; Score 1404.5; DB 2; Length 3623;
 Best Local Similarity 20.6%; Pred. No. 7.1e-71;
 Matches 695; Conservative 378; Mismatches 1083; Indels 1219; Gaps 128;
 QY 101 NEGYDLOGSKRITCMKVSMDPAWSDHPRVCRAWCDALHLPSCGIIIT--SPNFIQYDN 158
 DB 443 NGRCICIDINGFTC----DCTSSWTGYCTCPQAAACGILSGTGTAYHSPN--DTYIH 496
 QY 159 NAHCWITITLNPASKVTKLAFESFDLE-----RGYDTLTVGSGGQDQDKTVLYMSQA 212
 DB 497 NVNCFWIVRT-DESKVLHVTFTFDLBSANCPREY--LQHDGSSADPFLGRY----- 548
 QY 213 CSDSPHPIPGSLIPESMGDIWRQKTVLEICRDISSDARSGSVKSPKTSNAVELVAPG 272
 DB 549 C-----GSRPPGDIH-----SSANALYFHLYS 570
 QY 273 TEIEGSCGDPGIPAYGRREGSRFHGDTLKEGCPAFELVGQAKITCKKNQMSAKKPG 332
 DB 571 EYINSGR---GFTN-----RKEALKLPE 589
 QY 333 CVFSCFNFPTSPGCVLSPNYPEDYGNHLCVWLILARPSRIHLAFNDIDVEP----QF 388
 DB 590 CGGILTDNY-----GSITSFGYRPNVPRGRDCVQVNVNPNLITFTFGTILBSHNDCK 645
 QY 389 DFLVYKGGATAEAVLGTFSNGQLPSSITSSGHVAREFQTDHSTGKRGFTITTT--- 444
 DB 646 DYLEIRGGPFHQDVLGKFTCTSLTPPLKTTGPAARHIFHSDSSTSGKFAITLYLTQSD 705
 QY 445 -----PRHN-----ECPDPGVNVGKGFQSDLOLGSSISF 474

DB 706 LDGCGNTDTIDGELLPLPLSGPFSHRQCVYLITQAQGEQIVINFTHVELESQWGCSTHY 765
 QY 475 L-----CDEGLGQSGSTITTCVLRKGSVW-----NSAVLRG--EAPCGHLTSPS 519
 DB 766 IEVGDHSLLRKICGNITPLPISVSNSKWRIRLRIDALVQKASFRADYQVACGMLRG-E 824
 QY 520 GTILSPGMPFVYDALSCAWVIEAOPGYIKITFDPRK-----TEVNDTLEVRDGRTRYSA 575
 DB 825 GFRSPFYPMAYPRRCRCRTTISOPQROVVLNFTDQIOTSSASCDTDYIEIGPSSVLGS 884
 QY 576 PLIGVYGTQVPOPLISTSNVYLLFSTDKSHDIDGFLAREYETITLYQSDHCLDPGIPVNG 635
 DB 885 PGNKFFSSNIPSTISVNIILVYTFVKSSEMENRGFTAFSSDKLE----- 931
 QY 636 QRHNDYVQALVTFSCDSGTTLSDEPLCEPFWQSRALPSCBALCGPFGSGSTIL 695
 DB 932 -----CGEVLTAFTGITE 944
 QY 696 SPGPDPFYNNLNCWTIETSHGKVPFTHTPLR-----SGHDYLLITENGSTFOTLRQ 751
 DB 945 SPGHFNVPYPGVNCIMHVVVQRGOLILBESSFYLEPHVYCTNDYLETIYTLAOTF-LGR 1003
 QY 752 LTGSRLEPAPISAGLYGNFTAQVRFISDPNSYEGFNITFSEYDLEPCBEDEVPAYSIRKG 811
 DB 1004 YCGKSIPPSLTS--NSNSIKLIFVSDSALAHGFSINYEAD-----ASSV--- 1047
 QY 812 LQGVGDTLTFSCPPYRLGRTARITLGGRRRLMSPLRCVABCGNSVTGQGITLSP 871
 DB 1048 -----CLYDTIDFNGMLSSP 1062
 QY 872 NFPVNNVNNHCTYSIOTOPGKGIOLKARAFELSE-----GDVLKYDDNNNSARLALGV 925
 DB 1063 NFPNNVPSNNECTYRIITVGLNQOLAHFTDFTLBDYFGSCQDVFELRDGEGYELPVLGI 1122
 QY 926 FSHSEMMGVTLNSTSSSLMDFTDAENTSKGELHPSFELIKCEDPGTKRGYKVHDE 985
 DB 1123 YCGS-VLPPIIISHSNLWKFKPSDALTKGSAVMD----- 1159
 QY 986 GHPAGSSVSPSCDGVSLRGSBEILCLSGRRRTWDRPLPTCVABCGGTVRGEVSGVLSLP 1045
 DB 1160 -----GSSGTG-----CGGNLT-TPTGVLTSP 1179
 QY 1046 GYPAPYEHNLNCITWTEAEAGCTIGLHFLVFDTEBYV-----DYLRTMDPVSQVLKE 1100
 DB 1180 NYPPPYHSECTWRLEASHGSPPELRFQPFHLEH-HPSCSLYLAVIDPPTNSRLIDK 1238
 QY 1101 LSGPALPKDIASTFNSVVLQFSTDFTFSKOGFAIOF-----SVSTATSCNDPG 1148
 DB 1239 LCGDTTPAPIRSKNDVLLKLRTDAGQGSGFELNFRQCDNVIVNKTGILESINYPN 1298
 QY 1149 IPONGSRSGDSWEA-GDSTVFQCDPGYALOGSAEISCVKIE-----NRFWQSPPT 1199
 DB 1299 PYDKNQRCNWTIOATGNTVNTYFLGFVDSYNNCSITDYELVDGPMQMRGCGNNMP 1358
 QY 1200 CMAP-----CGGDLTGSSGYILSPNYPEPYPKCEK 1231
 DB 1359 GATTSQHLVLFHTDGINSEKGFKNQWFTHGCGGSGTAGSFSFGYNSYPHNECT 1418
 QY 1232 WKYTVSPDYIALVFNIFNLE-----PGYDLATHYDGDLSPLIGSPYG--SOLPQRIE 1284
 DB 1419 WNRIVAPGSSIQULTIHDFVHYHTSCYBDELEIYAGIDRSPRIALQCSGSPSAPNPQVS 1478
 QY 1285 SSSNSLFLAFRSDASVSNAGFVIDYENPRESCFPDPSIXNGTRVGSCLKGSSVITYCH 1344
 DB 1479 STGNELAIRKXTDSTLNGRGNFNASWRAVP----- 1507
 QY 1345 GGYEVBSTSLSCILGPDGKFPVNNRPVCTARCGGYGSDGVLSPNYPOWYTSQIC 1404
 DB 1508 GG-----CGGIIQISRGELHSPNYPNRYRATTEC 1536
 QY 1405 LYFVTVKDYVVGQAFHFH-TALNDVVEVHDGSHSRLSLSSGSHTESTLPLATSNQ 1463
 DB 1537 SWIIQVRRHHRVLLNTIDPDLAEPDGLRLMDSSSTNARVAVACRQGPNNIIASGNS 1596

QY 1464 VLKESAKGLAFARGFHHVQAVPRTSATQCCSVPEPRYGRKIGSDSFVGAIVAFECNCS 1523
 DB 1597 LFRVRSSSSGNRCFRABFR- 1618
 QY 1524 YALGSSPEIECLPVGALAAQNNVSAFPCVPCGN-LTERRGITLLSPGPEPYNLSNCV 1582
 DB 1619 -----ECGGRIMTSDTIFSPLYPHNYLHNCOS 1648
 QY 1583 WKI-VPEGAGIQIOVSF-----VTEQWMSLEVPDGDNTVMUGSSGTTVALNST 1637
 DB 1649 WIEKQPPPHNITLSFTFOLONSTDCRDFVELDNDYDAPVQGRYCGSLPHPTISF 1708
 QY 1638 SNOULHFYSDISVASAGFHELYKTIVLSCEPEAVSNVGTGERLYVNDVVSFQCEPG 1697
 DB 1709 GNALTREVTSTSSFEGRALYSA----- 1733
 QY 1698 YALGHNHISCMPTGVRMNYPPPLCIAQCCGTVIEHBGVILSGFPGNPSNMDCSWKI 1757
 DB 1734 -----STSSCGSFYTLDGIFNSPDYADYHPNAECVWNI 1768
 QY 1758 ALPVGFGAHIOFLNSTEP-----NHDYIEIRNGPYETSRMWGRSGSELSLSLT-SHE 1812
 DB 1769 ASSPNNLQSLFSLFENLNSLNCMDPVEIRFG-NATGHLIGKICGSLPGNYSAGHS 1827
 QY 1813 TTVTFHSDHSONRPGFKLEYQ-----AY 1835
 DB 1828 LWFVFSVDSGSGTGQFARFNIPGNNNIVGTHGKIASPMPGKYPNYSNYKVVANDAI 1887
 QY 1836 -----ELQECDDPPFANGIVRGAGV-----VG-----QVTECELP 1868
 DB 1888 HIHGRLEMDIEFTTCTFYDSLKIYDFDTHSLIGTCGTOTESFSSRNLSLTPOFS 1947
 QY 1869 GYOLTGHPVLTCQHTNRNMDHPLKCEVPCGNITSSNGTV--YSPGFPSRSSODCV 1926
 DB 1948 DSSVSGRFILEMFAVVDVSTPPTIAPAGCGGMVTDIPVHIFSGWREIRANADCI 2007
 QY 1927 MLTVPFGHVRNLISLOTETP-----SGDFTTWGPOQTAPRLGYFTSMARKTVQSS 1982
 DB 2008 WIIYAP-DSTVEINILSLDIEPOOSCNVDLIYDGDSDLSPELAVLCGVSPGPIRSNG 2066
 QY 1983 NOVLKFRDATGAI PAI AFSAVPLTKCP-----PPTILPNAEVTENE 2027
 DB 2067 EWMYIRTSIDTSAV-----TGFNNSFKHSCGGLHADRGVITSKYDYTLPMV-----NC 2118
 QY 2028 ENNIGDIVRYRCLPGEFTL-----VGNELITCKLGYLQF-EGP-----PPI 2067
 DB 2119 SMHV--LVQ-----TGLTIAVHFEQPOFOIQNRDSFCGQDYLVRNRPDMSHSPGSGRN 2172
 QY 2068 -----CEVHCPT-----NELL-----TDSGTVI 2085
 DB 2173 GRFCGMVABSTLTSGNEMFVOTISDSSNGGQFKIRYAKSLACGTVYIHDADSDGL 2232
 QY 2086 LSGSYPGSYPOFQTCMLVREBDYNIISLVEYFLSEKQYDF-----EIPD 2132
 DB 2233 TSPNYFANYPQHAECIWIIEAPGRSIOQFE-----DQENIEDTNCVSLIELRD 2284
 QY 2133 GBSGOSPLKALSGNSAPLITSSNSYLRMSDHAIVNRKFKIRYAPCSLPAPL 2192
 DB 2285 GANSNRILVSKLQGHLPMSWSSRERYLKFHTLGGSSVNGFAKYSIASC----- 2336
 QY 2193 HGFILGOTSTPGSITHFCNAGYRLVGHSMALCTRHPOGHLMEBAIPLQALSCGLPE 2252
 DB 2337 -----GTVSGDSGVLE--SIGYPLPYANNV-----FCQWPIRGAP- 2371
 QY 2253 APRNGNPFKEYTVGTKAVYSCSEGYHQAAGATAECLDTGLMSNR-----NV 2301
 DB 2372 -----GHYLTLS-----FEDENLQSSPGCTKDPFE--IMENHISGRVLGYCNS 2414
 QY 2302 PQQCVPTGPDVSI-----SVENRKMILLIETQYQFOAQLMLICDPGYTYTGQVIRIC 2355
 DB 2415 TFSVD-TSSNVAKYFVTGDSVTASGFLQPKSSRO-----VCG----- 2453

QY 2356 QANGKSLSD-STPTCRITISCELP-ENG-RI--GLISYGAIAIF----- 2399
 DB 2454 -----DDLHGP------GTFTSPNYNPNPHARICEMWITVQEGRIIVLFTNLRL 2499
 QY 2400 -----SCNSGYTLVAGVRRECANGMLSGSEVCLAGHCTPPPIVNGHINGENYSRGSV 2455
 DB 2500 STOPSNCSEHLIV-----ENGIRSN-PLUO----- 2524
 QY 2456 VYQCNAGFRLIGMSVRIQQ-----DHWSGKTPPCVPITTCGHPGNVNGLTQGNQFNLN 2510
 DB 2525 -----KLCGRVAVVNEFKSSGNTMKVFFTDG--SRPYGGL- 2559
 QY 2511 DVYKFCNMGWABGAARSQCLASQMSMLPTCIIINTDGHQBNSTR-----OV 2562
 DB 2560 -----ASTSTEDAVCGG--FLPSVSGGNFSSBGY--NGIRDYARNLDCW 2601
 QY 2563 HASGPHRSFGTTVSYRCHNGFYLLGTPLVSCQ-----GDGTWDRDRPQCLLYS 2611
 DB 2602 TLSNPRENSISL-----YFLELSIESHQDCTFVLEFRVDADGPLEKCSLSA 2653
 QY 2612 CGHPSGPPHSQ-----MGSD--SYTVGAVRYS--CTGKRT--LVGNSTRMGLDGHW 2658
 DB 2654 PTAPLVIPPQVWIHFVSNRVEYT-GFYIEYSFTDCGIRTDGNGVILSPVPLYGAM 2712
 QY 2659 TGSLEPHSGTSVGVCGDPIPAHGIRLDSPDPGTVMEFSEAGHVLARGSSERTQANGS 2718
 DB 2713 T-----HCSMLKAPBG-----HTTL-----TFSDFLBA-----HPTISDSV 2747
 QY 2719 W--SGSQEPCGVIS--CGNP-----GTPSNARVFPSDGLVFSSTIVYECRGGYATGLLS 2769
 DB 2748 TVRNGDSFGSPVIGRYCGQSVPRPIOSGNOLIV-----TENTNQGQTR-GFYA----- 2796
 QY 2770 RHCSVNGMTGSDPCLVINGDPCIPANGIRLGNDFRINKTVTYQCYPGYTMESHRSV 2829
 DB 2797 -----TWITN-----ALGGGTFHSANG-----TISPHMWQTFEPSSRSW 2833
 QY 2830 LSCTKD-----RTWNGTKPYCKALMC-----KPPPLIP 2857
 DB 2834 TVITHESGHWEISPDNSNRIPSSDSQONSFYKXWEGRLMINKLLIATSCGDVAPSPIVT 2893
 QY 2858 NGKVVSGDFPMWSSVTTACLEGYOLSLPAYTCE--GNGSWTGLPQCFPVFCGDPG-- 2912
 DB 2894 SGNL-----FTAVQSEEMAAQGSASFISRCGRFTNTSPDI 2931
 QY 2913 -VPSRGREDRGFS-----YRSVSFSCHPPLVLVGSPPRRFQSDG----- 2952
 DB 2932 ISPNFPKQYDNNMCTYLIDADPOSLVILFVSHLEDRSAITGT-----CDHDELHITKG 2987
 QY 2953 -----TWSGTQ--PSCID-PTLTT-CADPGVPOFGIIONNSQGVYSTVLFRQCK 2998
 DB 2988 RNLSTPLVITICSETLRPLTVDDPVLNLFYSDAVTTDFGKISYRAITCGG--IYNSS 3045
 QY 2999 GYLQSGSTTRCLEN 3013
 DB 3046 GILSPSISYSNYPN 3060

RESULT 15
 Q29530 PRELIMINARY; PRT; 2014 AA.
 AC Q29530;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Complement receptor 1 (Fragment).
 GN Name=CR1;
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 OX NCBI_Taxid=9598;
 RN [1]
 RP SEQUENCE FROM N. A.
 RX MEDLINE=94292799; PubMed=8021505;

RA. Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;
 RT "Primary sequence of an alternatively spliced form of CR1. Candidate
 RT for the 75,000 M(x) complement receptor expressed on chimpanzee
 erythrocytes."
 RL J. Immunol. 153:691-700(1994).
 DR EMBL: L24920; AAAS1438.1; -.
 DR PIR: I36936; I36936.
 DR GO: GO:0004872; F:receptor activity; IEA.
 DR InterPro: IPR000834; Peptidase M4.
 DR InterPro: IPR001424; SOD CU ZN_.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; Sushi_30.
 DR SMART: SM0032; CCP; 30.
 DR PROSITE: PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_2.
 DR PROSITE: PS00087; SOD CU ZN_1; UNKNOWN_1.
 DR PROSITE: PS50923; SUSHI; 30.
 DR Receptor.
 FM NON_TER
 FT NON_TER
 SQ SEQUENCE 2014 AA; 221280 MW; 6D6C3A74D81F1DB9 CRC64;
 Query Match 8.2%; Score 1390; DB 2; Length 2014;
 Best Local Similarity 22.2%; Freq. No. 2.1e-70;
 Matches 544; Conservative 266; Mismatches 820; Indels 822; Gaps 111;
 QY CEEDEVPA---YSIRKGLQFGVGDITLFPSCFPYRLBETARITC-LGGRRLMSSPLPRC 853
 DB CRNPDPDPNGMWHYIKDIOF--GSOIKSCTKGRILIGSSATCIIISGDVYIMNENRPIIC 136
 QY 854 -VAECGNSVTGTGTLSPNFPVNNHNCIYSIQTOPGKIQLKARAFELSGDVLK 912
 DB 137 DRICGLPPTITNGDIFISTN--KENPHYGSVVTYRCNPGSG--GRKVFEL-VGEP-SI 188
 QY 913 YDGNNSARLLGCVSHSEMVGVTINSTSSSLMDPITDAENTSIGFELHFSFELICED 972
 DB 189 YCTSNDDQ--VGISGAPQCIIIPN-----KCTP 215
 QY 973 PGTAKFGKYVDEGHF--AGSSVSPSCDGYSLGSEELLCLSGERRTWDRPLPTCYAEC 1030
 DB 216 PNVEN-GILVSDNLSLFSINEVEFRCPGFVWKPAPRYKCOLANK--WELBELSCSRVC 272
 QY 1031 GGTVRGEVSGVLSPGVAPAPYEHNLNCITWTEAENGCTIGLHPLVFTDEVDYLRIWDG 1090
 DB 273 -----QPP-----DVLH----- 280
 QY 1091 PVEBGVLKELSGPALPKDLHSTNSVVLQSTPFTSKGPAIQFSVSTASCNDPEIF 1150
 DB 281 -----A 281
 QY 1151 QNGSRSGDSWEAGSTVFCQCDPGVALGSGAEISCVKLENRFMQPSPTCIAFGGDLTG 1210
 DB 282 ERTQRDKDNFSPGGEVFPYCEGYDLRGASLRCTPGD--WSPATPCTVKSCDDPWG 338
 QY 1211 P--SGVILSPNYPPPYPKCKEWMKVTVSPDYIALVENILEPBGYFLHIYDGRDLS 1268
 DB 339 QLNGRVLP-----VNLQGAQVDFVCD----- 362
 QY 1269 PLISFYGSQLPGRIBSSNSLFLA---FRSDASVSNAGFVIDYTEMPPRESGPDPSI 1333
 DB 363 -----EGFQLKG---SSASYCVLAGMSLWNSSPVCEQIF-----CPSPPIVI 402
 QY 1334 KNGRVRGSDTLK---GSSVTVYCH-----GGYVEGTSLSLIGPDGRKVMNNPPRV 1374
 DB 403 PNGNHTGKPLFVPPGKAVNYTCDPHPRGTTFDIIIGESTIRCTSDPGANGWSSPARC 462
 QY 1375 TAPCGGYVSGDGVVLSPNYPQNTYSGQICLYFVTVPKDYVVFQPAFFHTALNDVVEH 1434
 DB 463 -----GIL----- 465
 QY 1435 DGHSG--HSRLISLISGHTGSLPLATSNOLIKESAKGLAPARF-----HFTYQAV 1486
 DB 466 -GHCQAPDHPFLFAKLKTQTNASDPFGITS---LKYKCRPRYGYGRFSGITCLDNLWSS- 519

QY 1487 PR--TATQSSVPEPRYK-RIGDSFSGAIVREPCNSGYALQSGPEIECLFVPGALQ 1543
 DB 520 PKDVCKRSCKTDPDPNGMWHVITDIQVSRINSCTTGHRLIGHSSABCI-LSGNAAH 578
 QY 1544 WNVSAPICV-VPCGGULTERGTILSPGPEEPLNSLNCWKIVBEGAGIQOVSFVT 1602
 DB 579 WSTKPPICORIPCGLEPTIANGDFISTNRENFHGS-----VVTYRC 620
 QY 1603 EQWMDSLVEFDGADNTFMTLGSFSGTTVPALINSTNQLYHFYSDISVSAAGFHEYKT 1662
 DB 621 NPSRGRKVE-----LVGERS-----IYCTSDN-----DQ 646
 QY 1663 VGLSSCEP--AVPS-----NGVYKGER--YLWVDVVSFQCEPGVALQGHAIISM 1709
 DB 647 VGISGAPQCIIIPNKTCPNVENGLIVSDNRLFSINEVEFRCPGFAMKPRRYKCQ 706
 QY 1710 PGTVRKNVPPPLCIAOCGGTVEEMSVILSPFGNYPGNSMCKIALPVFGAHIOF 1769
 DB 707 --ALNKKWEPPLPSCSRVC----- 722
 QY 1770 LNFSTEPNHDYIELRNGPYETSRMMGRFSGSELPSLLSTSHETTVFHSDSQNRPGFK 1829
 DB 723 -----QPPDVL-----HAKTOR----- 736
 QY 1830 LEYQAVELQPCDPPEPANGIVRAGYNGQSVTFECLPGYOLTGHPVLTCQHTNRMD 1889
 DB 737 -----DKDN-----SPQEVFYSCPEPYDLRGASLACT--PGDWS 772
 QY 1890 HPLPKCEV-PCGNITIS-SNGITYSPGFSPYSSQDCWLITVPYIGHGYRLMSLLQTE 1947
 DB 773 PAAPTCVKSCDGFMGQLNGRVLP-----VNLQGAKV----- 807
 QY 1948 PSGDFTIMDGPQOTAPRLGVFTSRMAKTKVQSSNOVLKFRHDATGIFAIAPASAP 2007
 DB 808 --DFVCD-EGFQLKG-----SSASICVL-----AGMSLWNSVP 839
 QY 2008 LTK---CPPTIIPNAEVVTEN--EEFNIGDIVYRCIP-----GFTLVGNEIITCKL- 2055
 DB 840 VCGDIFPSPFVLPNGHTKPLEVPPGKAVNYTCDPHPRGTTPDLIGESTIRCTSDP 899
 QY 2056 --GTLYQFEGPPPICV--HCPTELLDSTGVILSOSYPSYPOFQTCMWLVREPDY 2110
 DB 900 QGNGVW---SSPARCGIILGHCOA-----PD- 922
 QY 2111 NISLTVEXELSEKQYDFELFDGSPGSPILKALSGVNSAPLVTSSNSVYLRMSDHA 2170
 DB 923 -----HFLPAKLKTQTNASDPPIGTSLYEKREPREYGRFSITCLDNLV---WSSPKD 972
 QY 2171 YNRKGFIRYSAPYCSLPRAPLHGFILOSTPOGSIHFGCAGYALVGHMAICTRHP 2230
 DB 973 VCHR-----KSKTTPDPVNGMWHVITDIQVSRINSCTTGHRLIGHSSABCTLSG 1024
 QY 2231 QGYHLMSEALPLCOALSQGLPEAPKNGWVG---KEYTVGKAVYSCSEGYH---LOAG 2283
 DB 1025 NSAH-WSTKPPICORIPCGLEPTIANGDFISTNRENFHGSVVTYRCNLGSRGRKVEY 1083
 QY 2284 AEATAECL---DTGLMS-----NRNVP-----QCV 2307
 DB 1084 GEPSIYCTSNDDQVIGISGAPQCIIIPNKTCPNVENGLIVSDNRLFSINEVEFRCP 1143
 QY 2308 -----VTC-----PDVSSIS-----VEHGRRLIFETQYQFOALMLICDP 2343
 DB 1144 GFWMKGRHRYKCOLANKWELBELSCSRVCOPPEIILHGETTBHQDNFSGQEVFVSCPE 1203
 QY 2344 GYVYTGQVRVRCOANGKMSLGDSPTCRILIS---GELPIPNGHRIIGTLYV-YGATAI 2398
 DB 1204 GYDLRGASLHCTPOGDS--PEAPRCAYVSCDDFLGQL---PHGRVLPFLNLQLAGKVS 1258
 QY 2399 FSCNSGYTVGSRVRECMANG---LMSGSVRCLAGCGTPEPIVNGHGENYVS---YR 2452
 DB 1259 FVCDDEGRILGSSSVSHCVLVGMSRLWNSVAVCEQIFCPNPALILNGRHTGTDFGDI 1318
 QY 2453 GSVVYQCN-----AGRLIGMSVRLCQDQHH---WSGKTPTPC-VPTCGH----- 2493

```
Db      1319 KEISYTCDDPHDRGMTFNLIIGESTIRCTSDPHGNGVWSSPAAPRCELPHAGHCKTPEQFP 1378
Qy      2494 ---PCNPVNLJTGQNOFNLDVVKFVNCNPGYMAEGARSOCLASGQWMDLPTCRILINCT 2550
Db      1379 PASPTLFINDF---EPVGTSLVNECRPGYFGMFSIS-CLENLVWSSVEDNCRKSCG 1433
Qy      2551 DPGHOENSVRQVHASGPHRFSFGTTVSYRCNHGFYLLGTPLSCQGDG---TMDRPRPOC 2607
Db      1434 PPEPEFGWVHINTD---TQFGSTVWVSCNEGRLLGSPSTCIVSGNNVTWDXKAPIC 1489
Qy      2608 LTVSCGHPGSPPHSQNSGDSYT-----VGAVRYS-----IGKRTLVGNS 2648
Db      1490 EIISC---EPPTISNGDFYSNNRASFNHGTVTYQCHTGPDEQLFELVGERSIY-C 1543
Qy      2649 TRMCGLDGHWLGSLLPHCSGTSVGVCGDPGI PAHGIRLGDG---FDPGTWRFSCGAGHYL 2705
Db      1544 TSDDDQGVWSSPPPRCTISTN--KCTABEV-ENAIRVPGNRSPFSLTEIVRFRCQPGFWM 1600
Qy      2706 RGSSERTQANGWSGSGQPCGVISCGNP-----GTPSNARVPSDGLVFSSSIYEC 2758
Db      1601 VGSHTVQCCQTNMGWMPRLPHCSRV-CQPPPEILHGEHTPSH-----QDNFSPGQEVFYGC 1654
Qy      2759 REGYVATGLLSRHCSVNGTWSGDPBCLVINCBD--PGIPANGRLGNDFRYNKVTYQC 2816
Db      1655 EPGYDLRGAASLHCTPGQDWSPEAPRCTVKSCDPLGQLPHGRVLPFLNLQLGAKVSFVC 1714
Qy      2817 VPGYMMESHRSV--VLSCTDRMTNGTKPVCKALMCKPRPLIPNGKVGS---DFMWSGSS 2871
Db      1715 DEGFRLKGRASHCVLGKMK-ALMNSSVFVCEQIFCPNPPAILNGRHGTGTPFGDIPYKE 1773
Qy      2872 VTYAC-----LEGYQLSLPAVFTCEGNGSWTGLPQC---FPVFCGDPGVPSRG 2917
Db      1774 ISYACDTHPRGMTFNLIIG-ESSIRCTSDPGNGVWSSPAAPRCELSVPAACPHP--PKIQ 1830
Qy      2918 RREDRG-----FSYRSSVSFSCHPLVLVGSPPRFCSQSDGTSQTQPSCTIDPTLTTCADP 2972
Db      1831 NGHDIGHVSLLYLPGMTISYICDPGYLLVGKGFIFCTDGIWSQLDHYCKE--VNCSPF 1887
Qy      2973 GVPQFGIOWN---SOGYQVGSVTLFRCQKGYLLQGSTTRTCLPNLTWSGTTP 3021
Db      1888 -LFMNGISKELEMKVYHYGDYVTLKCEDGYTLEGSFWSQCADRMD--PP 1936
```

Search completed: October 18, 2004, 10:24:50
Job time : 191 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 09:57:30 ; Search time 109 Seconds
(without alignments)
10215.557 Million cell updates/sec

Title: US-10-016-248-2
Perfect score: 16985
Sequence: 1 MAGAPPALLPCLSLSDCC.....RSGPVDPTSLPESHSPKR 3104

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: geneseqp1980s:.*
2: geneseqp1980s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003ae:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16985	100.0	3104	5	ABG79168 Human cub
2	16985	100.0	3104	8	ADH71166 Human pro
3	16985	100.0	3104	8	ADH71138 Human pro
4	16978	100.0	3104	8	ADH71168 Human pro
5	16057	94.5	3546	8	ADH71136 Human pro
6	15655	92.2	3483	8	ADH71144 Human pro
7	15219	89.6	3130	8	ADH71146 Human pro
8	14142	83.3	2669	8	ABG79169 Human cub
9	14142	83.3	2669	8	ADH71140 Human pro
10	13142	77.4	2612	8	ADH71142 Human pro
11	11247	66.2	3567	8	ADH72216 Human pro
12	10760	63.4	3100	5	AAE20789 Human C3b
13	10755	63.3	3100	5	AAE20901 Human C3b
14	10720	63.1	3095	5	AAE20788 Rat C3b/C
15	10701	63.0	3069	5	AAE20787 Human C3b
16	10696	63.0	3069	5	AAE20900 Human C3b
17	10475	61.7	3389	7	ADJ70480 Human hea
18	9025	53.1	1958	5	AAO19415 Human mol
19	7662	45.0	2352	5	AAU11816 Cancer an
20	7643	45.0	2306	5	AAU11817 Cancer an
21	7568	44.6	2008	5	AAU11814 Cancer an
22	7098	41.8	1826	5	AAU11812 Cancer an
23	6949	40.9	1800	5	AAU11813 Cancer an
24	6936	40.8	1783	5	AAU11815 Cancer an
25	4937	29.1	882	4	AAH83372 NOV16 pro

ALIGNMENTS

26	4937	29.1	882	8	ADH71164 Human pro
27	4869	28.7	883	4	AAH83371 NOV15 pro
28	4859	28.6	884	8	ADH71162 Human pro
29	4692	27.6	869	8	ADH71154 Human pro
30	4268	25.1	776	8	ADH71150 Human pro
31	4257	25.1	776	8	ADH71148 Human pro
32	4143	24.4	1274	5	ABG76507 DNA encod
33	3856	22.7	713	8	ADH71152 Human pro
34	3174	18.7	613	4	AAU00816 Human Imm
35	2979	17.5	597	8	ADH71158 Human pro
36	2651	15.6	810	8	ADH72220 Human pro
37	2639	15.5	529	5	ABP43478 Human sec
38	2555	15.0	839	5	AAE23384 Human int
39	2416	14.2	601	7	ADG31384 Human nov
40	2105	12.4	413	7	ADG31488 Human nov
41	1850	10.9	3594	5	AAE20147 Mouse C3b
42	1794	10.6	3571	5	AAE20146 Human C3b
43	1793	10.6	3570	5	ABJ10589 Human nov
44	1793	10.6	3570	8	ADH71302 Human pro
45	1793	10.6	3570	8	ADH71304 Human pro

RESULT 1

ABG79168 standard; protein; 3104 AA.

ABG79168; (first entry)

15-NOV-2002 (first entry)

Human cub and sushi domain containing protein #1.

KW Human; NOVX; neurodegenerative disease; Alzheimer's disease; anxiety;
KW Parkinson's disease; Huntington's disease; neurological disorder;
KW schizophrenia; manic depression; mental retardation; angina pectoris;
KW cardiovascular disease; acute heart failure; myocardial infarction;
KW muscular disease; muscular disorder; retinal disease; photoreception;
KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;
KW immunological disorder; inflammatory disease; immune disease; diabetes;
KW bacterial infection; fungal infection; protozoal infection; obesity;
KW viral infection; reproductive system disorder; metabolic disturbance;
KW anorexia; wasting disorder; chronic disease; infectious disease;
KW dyslipidaemia; cub; sushi; myelin; von Willebrand factor; kielin;
KW semaphorin; serine/threonine protein kinase; TGF-beta binding;
KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;
KW toll-like 2; cysteine sulfinic acid decarboxylase SNP;
KW single nucleotide polymorphism.

Homo sapiens.

Key Location/Qualifiers

FT misc_difference 1027 /note="Ala substituted by Thr as a result of a single nucleotide polymorphism (SNP)"

FT misc_difference 1084 /note="Val substituted by Ala as a result of a single nucleotide polymorphism (SNP)"

FT misc_difference 1362 /note="Asp substituted by Gly as a result of a single nucleotide polymorphism (SNP)"

FT WO200264791-A2.

PN 22-AUG-2002.

XX 10-DEC-2001; 2001MO-US048369.

XX 08-DEC-2000; 2000US-0254329P.

XX 14-DEC-2000; 2000US-0255648P.

PR 15-MAY-2001; 2001US-0291037P.

PR 08-JUN-2001; 2001US-0297173P.
 PR 08-JUN-2001; 2001US-0309258P.
 PR 29-AUG-2001; 2001US-0315639P.
 PR 01-OCT-2001; 2001US-0326393P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Alsebrook JP, Anderson DM, Burgess CE, Boldog FL, Casman ST,
 PI Colman SD, Elinger SR, Ellerman K, Gerlach V, Gotman L, Grose WM,
 PI Guo X, Hermann JL, Kekuda R, Lepley DM, Li L, MacDougall JR,
 PI Miller I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkels RA,
 PI Smuthson G, Splyek KA, Stone DJ, Tchernev VT, Vernet CM, Voss EZ,
 PI Zetunen BD, Zhong H, Zhong M;
 XX
 DR WPI; 2002-643486/69.
 DR N-PSDB; ABS64375.
 XX
 PT New NOVX polypeptides and polynucleotides useful for treating or
 PT preventing e.g. neurodegenerative diseases, neurological disorders,
 PT cardiovascular diseases, muscular diseases and disorders, or
 PT immunological diseases.
 XX
 PS Claim 1, Page 13-14; 239pp; English.
 XX
 CC The present invention relates to new NOVX polypeptides. The polypeptides,
 CC polynucleotides and antibodies are useful in the manufacture of a
 CC medicament for treating or preventing neurodegenerative diseases (e.g.
 CC Alzheimer's disease, Parkinson's disease, or Huntington's disease),
 CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or
 CC mental retardation), cardiovascular disease (e.g. acute heart failure,
 CC aortic pectoris or myocardial infarction), muscular diseases and
 CC disorders, retinal diseases (including those involving photoreception,
 CC deafness and keratinisation disorders), cancer (e.g. ovarian cancer or
 CC melanoma), immunological disorders, inflammatory and immune diseases,
 CC bacterial, fungal, protozoal and viral infections, and reproductive
 CC system disorders. The proteins of the invention may be used to screen
 CC drugs or compounds that modulate the NOVX protein activity or expression,
 CC as well as to treat disorders characterised by insufficient or excessive
 CC production of NOVX protein or protein forms that have decreased or
 CC aberrant activity compared to NOVX wild type protein, such as diabetes,
 CC obesity, metabolic disturbances associated with obesity, anorexia and
 CC wasting disorders associated with chronic diseases and various cancers,
 CC infectious diseases and various dyslipidaemias. The nucleic acid
 CC sequences of the invention may be used in chromosome mapping, identifying
 CC an individual from minute biological samples (tissue typing), and in
 CC forensic identification of a biological sample. The present amino acid
 CC sequence represents a NOVX protein of the invention
 XX
 SQ Sequence 3104 AA;
 Query Match 100.0%; Score 16985; DB 5; Length 3104;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 TLKECQPAFELVGVKAITCOKNNQWAKKPGCVSCFPNFSTSSGVLSPNVPEDEYGNH 360
 DB 301 TLKECQPAFELVGVKAITCOKNNQWAKKPGCVSCFPNFSTSSGVLSPNVPEDEYGNH 360
 QY 361 LHCWMLIARPSRIRHLANDIDVEPQPLVIXGATAEAVLGTFSGNOLPSSITSSG 420
 DB 361 LHCWMLIARPSRIRHLANDIDVEPQPLVIXGATAEAVLGTFSGNOLPSSITSSG 420
 QY 421 HVARLEFQDHTSGKRGFNITFTFRHNECPDPGVNVNKRFGDSIQGLSSISFLCDEGF 480
 DB 421 HVARLEFQDHTSGKRGFNITFTFRHNECPDPGVNVNKRFGDSIQGLSSISFLCDEGF 480
 QY 481 LGTQSEITTCVLKRGSVVNSAVLRCAPCGGHLTSPSGTLLSPGMPGYKDALSCAWV 540
 DB 481 LGTQSEITTCVLKRGSVVNSAVLRCAPCGGHLTSPSGTLLSPGMPGYKDALSCAWV 540
 QY 541 IEAOPGYPIKIFDEPKEVNVDTLEVRDGRYSAPLIGVHGTQVPOFLISTSNVYL 600
 DB 541 IEAOPGYPIKIFDEPKEVNVDTLEVRDGRYSAPLIGVHGTQVPOFLISTSNVYL 600
 QY 601 FSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHGNDFYVGLVTFSCDSGYTSLD 660
 DB 601 FSTDKSHSDIGFQLRYETITLQSDHCLDPGIPVNGQRHGNDFYVGLVTFSCDSGYTSLD 660
 QY 661 GEPLECEPFWMSRALBCECALCGFICGSGTLLSPGPPDYNNLCTWIIETSHSGK 720
 DB 661 GEPLECEPFWMSRALBCECALCGFICGSGTLLSPGPPDYNNLCTWIIETSHSGK 720
 QY 721 VFFTHHTHLESGHYLLITENGSTFQPLRLQTLGSRLLPAPISAGLYGFTQVRIISFS 780
 DB 721 VFFTHHTHLESGHYLLITENGSTFQPLRLQTLGSRLLPAPISAGLYGFTQVRIISFS 780
 QY 781 MSYEFNITFSEYDLEPCEBEVPAYSIKRLQFVGDTLTFSCPPGYRLGTARITCLG 840
 DB 781 MSYEFNITFSEYDLEPCEBEVPAYSIKRLQFVGDTLTFSCPPGYRLGTARITCLG 840
 QY 841 GRRRLMSPPLRCVAECGNSVTGQTLTSLNPNVNNNECTYISTOTQKGKIQLKAR 900
 DB 841 GRRRLMSPPLRCVAECGNSVTGQTLTSLNPNVNNNECTYISTOTQKGKIQLKAR 900
 QY 901 AFELSEGQVLKYVDGNNNSARLLGVFSHEMVGVTLSSTSSLWLDFTDAENTSKGEL 960
 DB 901 AFELSEGQVLKYVDGNNNSARLLGVFSHEMVGVTLSSTSSLWLDFTDAENTSKGEL 960
 QY 961 HFSSELIKCDPCTPKRGYVNHDEGHPAGSVSFSCDPGSLRGSEBLLCLSGERRTWD 1020
 DB 961 HFSSELIKCDPCTPKRGYVNHDEGHPAGSVSFSCDPGSLRGSEBLLCLSGERRTWD 1020
 QY 1021 RPLPTVAECGGTVRGEVSGQVLSPGYPAPAEHNLNCTWTEBAGCTIGLHFLVPTDEE 1080
 DB 1021 RPLPTVAECGGTVRGEVSGQVLSPGYPAPAEHNLNCTWTEBAGCTIGLHFLVPTDEE 1080
 QY 1081 VHDVLRITWDGVESGVLLKELSGPALPKDLHSTFNSVLQFSTDFETSKQGFALQFVSST 1140
 DB 1081 VHDVLRITWDGVESGVLLKELSGPALPKDLHSTFNSVLQFSTDFETSKQGFALQFVSST 1140
 QY 1141 ATSCNDPGIIPONGRSRSDSWEAGDSTVFQCDPGIALGSAIEISVKTLENFFMQPSPTC 1200
 DB 1141 ATSCNDPGIIPONGRSRSDSWEAGDSTVFQCDPGIALGSAIEISVKTLENFFMQPSPTC 1200
 QY 1201 IAPCGDULTPSGVILSPNYPPEYPPEKEDMWKTVSPDVIALVFNILPEPYDTLHI 1260
 DB 1201 IAPCGDULTPSGVILSPNYPPEYPPEKEDMWKTVSPDVIALVFNILPEPYDTLHI 1260
 QY 1261 YDGRDLSPLIGSYGQLGRLTSSNSGLFLAFRSDASVSNAGFVITYENPRESCFDP 1320
 DB 1261 YDGRDLSPLIGSYGQLGRLTSSNSGLFLAFRSDASVSNAGFVITYENPRESCFDP 1320
 QY 1321 GSIKNGRVSVDLKLGSVITYCHGGEVEGTSITLGIAPDGPVWNNRPVCTAPCGG 1380
 DB 1321 GSIKNGRVSVDLKLGSVITYCHGGEVEGTSITLGIAPDGPVWNNRPVCTAPCGG 1380

QY 1381 QYVSGDVVLSVNYVPOVNTSGOICLYEVTVPKDYVVGQAPAFHTALNDVVEVHDGSHQ 1440
 DB 1381 QYVSGDVVLSVNYVPOVNTSGOICLYEVTVPKDYVVGQAPAFHTALNDVVEVHDGSHQ 1440
 QY 1441 SRLLSLSGSHGESJPLATSNQVLKFSAKGLAPAGFHFVYQAVERTSATOCSSVPEP 1500
 DB 1441 SRLLSLSGSHGESJPLATSNQVLKFSAKGLAPAGFHFVYQAVERTSATOCSSVPEP 1500
 QY 1501 RYKRLGSDSVGAIYAFECNSGYALOGSPETIECLPVPGLAOMNVSAPFCVPCGGLT 1560
 DB 1501 RYKRLGSDSVGAIYAFECNSGYALOGSPETIECLPVPGLAOMNVSAPFCVPCGGLT 1560
 QY 1561 ERAGTILSPGPPEPYLNSLNCWKIYVBEAGIIOVVSFVTEBOMNDSLEVPDQADTVT 1620
 DB 1561 ERAGTILSPGPPEPYLNSLNCWKIYVBEAGIIOVVSFVTEBOMNDSLEVPDQADTVT 1620
 QY 1621 MLAGSFQITVPALLNSTSNQLYLHFYSDISVSAAGFLBXYKTVGLSSCPBPAPVSNQVKT 1680
 DB 1621 MLAGSFQITVPALLNSTSNQLYLHFYSDISVSAAGFLBXYKTVGLSSCPBPAPVSNQVKT 1680
 QY 1681 GERLYVNDVVSFQCEPGYALOGHAHISCMPTVRMNYPPPLCIAQCGTVEEBGVILS 1740
 DB 1681 GERLYVNDVVSFQCEPGYALOGHAHISCMPTVRMNYPPPLCIAQCGTVEEBGVILS 1740
 QY 1741 PGFPGNVPNSMDCSMKIALPVGGAHIQPLNFSTEPHDIETIRNGPYETSRMMGFRSGS 1800
 DB 1741 PGFPGNVPNSMDCSMKIALPVGGAHIQPLNFSTEPHDIETIRNGPYETSRMMGFRSGS 1800
 QY 1801 ELPSLSLSHSTHTTVVPHSDHSONRPGFKLEYQAYELQECRDEPFPANGIVRAGVNVQ 1860
 DB 1801 ELPSLSLSHSTHTTVVPHSDHSONRPGFKLEYQAYELQECRDEPFPANGIVRAGVNVQ 1860
 QY 1861 SVTFPECLPGYQLTGHPRVLTQCHGTNRNWDHLPKCEVPCCGNITSSNGITYVSGFSPYS 1920
 DB 1861 SVTFPECLPGYQLTGHPRVLTQCHGTNRNWDHLPKCEVPCCGNITSSNGITYVSGFSPYS 1920
 QY 1921 SSQDCWLVITVPIGHGRLNLSLQTEPSGDFITMGPPOOTAURLGVFTSRMAKKTVO 1980
 DB 1921 SSQDCWLVITVPIGHGRLNLSLQTEPSGDFITMGPPOOTAURLGVFTSRMAKKTVO 1980
 QY 1981 SSNOVLKFKHRDADATGIFAIASFAYBLTKCPPTILPNAEVTENEFNIGDIVRCL 2040
 DB 1981 SSNOVLKFKHRDADATGIFAIASFAYBLTKCPPTILPNAEVTENEFNIGDIVRCL 2040
 QY 2041 PGFTLVGNELITCKLGTLYLOPEGPPICEVHCPTNELITSTGYILSOSTPGSTPQTC 2100
 DB 2041 PGFTLVGNELITCKLGTLYLOPEGPPICEVHCPTNELITSTGYILSOSTPGSTPQTC 2100
 QY 2101 SWTVRVEPDVNIISLTVEYPLSEKQYDEFELPDGSGOSPLLKALSGVSAPLIVTSSNS 2160
 DB 2101 SWTVRVEPDVNIISLTVEYPLSEKQYDEFELPDGSGOSPLLKALSGVSAPLIVTSSNS 2160
 QY 2161 VYLRMSDHAUNRKGFKIRYSAPYCSLPRAPLHGFILGOSTQPGGSIHFGCNAAGVYLV 2220
 DB 2161 VYLRMSDHAUNRKGFKIRYSAPYCSLPRAPLHGFILGOSTQPGGSIHFGCNAAGVYLV 2220
 QY 2221 HSNMICTRHPQGHNLMSBAIPLCOALSGLPEABKCNMVFGEKXYVTGKAVYSCSEYHL 2280
 DB 2221 HSNMICTRHPQGHNLMSBAIPLCOALSGLPEABKCNMVFGEKXYVTGKAVYSCSEYHL 2280
 QY 2281 OAGGEAFAECIDTGLMSNRNVPQCVPTCPDVSSIVEGRMVLITETOQFQAQMLI 2340
 DB 2281 OAGGEAFAECIDTGLMSNRNVPQCVPTCPDVSSIVEGRMVLITETOQFQAQMLI 2340
 QY 2341 CDBPGYUUTGQVIRICQANGKMSLGDSTPTCKIISCGELPIPNHNRIGTISVYGATIF 2400
 DB 2341 CDBPGYUUTGQVIRICQANGKMSLGDSTPTCKIISCGELPIPNHNRIGTISVYGATIF 2400
 QY 2401 CNSGYTLLVGRVRECMANGMSGSEVRCLAGHCTPEPVIYNGHINGENYSTRGSVYQCN 2460
 DB 2401 CNSGYTLLVGRVRECMANGMSGSEVRCLAGHCTPEPVIYNGHINGENYSTRGSVYQCN 2460
 QY 2461 AGFRLIGMSVRLICQODHMSGKTPFCVPLTCGHGNGVNLGTQONQNLNDVYVFCVNGP 2520

DB 2461 AGFRLIGMSVRLICQODHMSGKTPFCVPLTCGHGNGVNLGTQONQNLNDVYVFCVNGP 2520
 QY 2521 YNABGAAROCCLASGWSMDLPTCRLINCTDPHOENSTRQVHASGPHRPSFCTTYSYC 2580
 DB 2521 YNABGAAROCCLASGWSMDLPTCRLINCTDPHOENSTRQVHASGPHRPSFCTTYSYC 2580
 QY 2581 NHGFYLLGTVPVLSQGDGTMDRPQCLVSCGHPSPPHOSMGSDSYTGAVVRYSCTG 2640
 DB 2581 NHGFYLLGTVPVLSQGDGTMDRPQCLVSCGHPSPPHOSMGSDSYTGAVVRYSCTG 2640
 QY 2641 KRTLGNSTRMCGLDHMTGSLPHCSGTSVGVCGDPIPAHGIRLGDSPDPGTVMRFSCE 2700
 DB 2641 KRTLGNSTRMCGLDHMTGSLPHCSGTSVGVCGDPIPAHGIRLGDSPDPGTVMRFSCE 2700
 QY 2701 AGHVLRGSSERTCOANGSWSGSOPECGVISCNPGTSPNARVVSGLVFSSTIYECRE 2760
 DB 2701 AGHVLRGSSERTCOANGSWSGSOPECGVISCNPGTSPNARVVSGLVFSSTIYECRE 2760
 QY 2761 GYATGLLSRHGCVNGTMTGSDPECLVINCGRPIRANGRLGNDPRYKVTYVOCVPGY 2820
 DB 2761 GYATGLLSRHGCVNGTMTGSDPECLVINCGRPIRANGRLGNDPRYKVTYVOCVPGY 2820
 QY 2821 MESHRSVYLSCTKDRTNMGTKRVCALMCKPPLIPNGKVSGSDPMWGSSTYACLBY 2880
 DB 2821 MESHRSVYLSCTKDRTNMGTKRVCALMCKPPLIPNGKVSGSDPMWGSSTYACLBY 2880
 QY 2881 QLSLPAVFTCEGNGSNTGELPOCFVPCGDPVPSGRREDRGSYRYSVSEFCBPVL 2940
 DB 2881 QLSLPAVFTCEGNGSNTGELPOCFVPCGDPVPSGRREDRGSYRYSVSEFCBPVL 2940
 QY 2941 VGSRRFCOSDGTWSTQSDCLDPTLTTCADPEVPFGIOMNSQGIQVGSITVLFRCQGY 3000
 DB 2941 VGSRRFCOSDGTWSTQSDCLDPTLTTCADPEVPFGIOMNSQGIQVGSITVLFRCQGY 3000
 QY 3001 LLOGSTTRCTLPMLTWSGTPRDCVPHNCRQPEPTANVAGALDLPMSGYTLITPARASP 3060
 DB 3001 LLOGSTTRCTLPMLTWSGTPRDCVPHNCRQPEPTANVAGALDLPMSGYTLITPARASP 3060
 QY 3061 SRVAPSTAPARMMAQASRPASMRSGPYGDPSTLPSHSRSPK 3104
 DB 3061 SRVAPSTAPARMMAQASRPASMRSGPYGDPSTLPSHSRSPK 3104
 RESULT 2
 ADH71166 standard; protein; 3104 AA.
 ADH71166:
 25-MAR-2004 (first entry)
 Human protein of the invention NOVap SEQ ID NO:62.
 human; cytosolic; immunomodulator; neuroprotective; nootropic;
 anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
 vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
 obesity; diabetes; infectious disease; metabolic syndrome X;
 dyslipidemia.
 Homo sapiens.
 WO2003102155-A2.
 11-DEC-2003.
 03-JUN-2003; 2003MO-US017430.
 03-JUN-2002; 2002US-0385120P.
 04-JUN-2002; 2002US-0385784P.
 05-JUN-2002; 2002US-0386041P.
 05-JUN-2002; 2002US-0386047P.
 06-JUN-2002; 2002US-0386376P.

[illegible]

QY 481 LGTQSGEITTCULKEGVSVMNSAVLRCEAPCGHLTSPGCTTILSPGMPGYKALSCAW 540
 DB 481 LGTQSGEITTCULKEGVSVMNSAVLRCEAPCGHLTSPGCTTILSPGMPGYKALSCAW 540
 QY 541 IEAOPGPKITIDRFKTEVNVDTLLEVRDRTYAPLIGVHGTQVPOFISTNLYL 600
 DB 541 IEAOPGPKITIDRFKTEVNVDTLLEVRDRTYAPLIGVHGTQVPOFISTNLYL 600
 QY 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDFYVALVTFSCDSGYTLSD 660
 DB 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI PVNGOHGNDFYVALVTFSCDSGYTLSD 660
 QY 661 GEBLECEPNFOWRALPSCALCGGFIQSSGCTILSGFPDPYNNLNCWIIETSHGK 720
 DB 661 GEBLECEPNFOWRALPSCALCGGFIQSSGCTILSGFPDPYNNLNCWIIETSHGK 720
 QY 721 VFPTFTPHLESGHDYLLITENGSPFOPLRLQTLGSRLPAPISAGLYGNFPAQVRFISDFS 780
 DB 721 VFPTFTPHLESGHDYLLITENGSPFOPLRLQTLGSRLPAPISAGLYGNFPAQVRFISDFS 780
 QY 781 MSYEGFNITSEYDLBCEBEPVAVSIRKGLQFVGDTLTFSCFPGRYLEGTAITCLG 840
 DB 781 MSYEGFNITSEYDLBCEBEPVAVSIRKGLQFVGDTLTFSCFPGRYLEGTAITCLG 840
 QY 841 GRRRLMSSPLPRCAVECGNSVTGCTGLLSPNFVANNHNECTYSTOTPGKIOUKAR 900
 DB 841 GRRRLMSSPLPRCAVECGNSVTGCTGLLSPNFVANNHNECTYSTOTPGKIOUKAR 900
 QY 901 AFELSEGDVUKYDGNNSARLLGVFSEMMGVTLNSTSSLMLEDTITAEVNTSKGFL 960
 DB 901 AFELSEGDVUKYDGNNSARLLGVFSEMMGVTLNSTSSLMLEDTITAEVNTSKGFL 960
 QY 961 HFSSFELIKCEDPTPKFGYKVAHDEGHFAGSSVSFSCDPGYSLRGSEELLCTSGERTWD 1020
 DB 961 HFSSFELIKCEDPTPKFGYKVAHDEGHFAGSSVSFSCDPGYSLRGSEELLCTSGERTWD 1020
 QY 1021 RPLPTCAECCGTVRGEVSGQVLSPGYPARYENHLMCIWTEABAGCTIGLHLVPEETES 1080
 DB 1021 RPLPTCAECCGTVRGEVSGQVLSPGYPARYENHLMCIWTEABAGCTIGLHLVPEETES 1080
 QY 1081 VHVULRIWDPVRESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFTSKGFALQFESVT 1140
 DB 1081 VHVULRIWDPVRESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFTSKGFALQFESVT 1140
 QY 1141 ATSCNDPGLPONGSRSGDSWEAGDSTVFOCDPGYALQGSABEISCVKIENRFPMQSPBTC 1200
 DB 1141 ATSCNDPGLPONGSRSGDSWEAGDSTVFOCDPGYALQGSABEISCVKIENRFPMQSPBTC 1200
 QY 1201 IAPCGGDLTGPBGVYILSPNTPBPYRPGKCDMKYTVSPDYVALVFNIENLREBGPDLAT 1260
 DB 1201 IAPCGGDLTGPBGVYILSPNTPBPYRPGKCDMKYTVSPDYVALVFNIENLREBGPDLAT 1260
 QY 1261 YDRDLSPLIGSFYGSQLPGRJESSNSLFLAFRSPASVSNAGFVLDYENPRESCFDP 1320
 DB 1261 YDRDLSPLIGSFYGSQLPGRJESSNSLFLAFRSPASVSNAGFVLDYENPRESCFDP 1320
 QY 1321 GSIKNGTRVGSDLKLGSSVTTYCHGYEVEGTSTLSCILGPDGKPVNNNRPVCTACGG 1380
 DB 1321 GSIKNGTRVGSDLKLGSSVTTYCHGYEVEGTSTLSCILGPDGKPVNNNRPVCTACGG 1380
 QY 1381 QYVGSDBVULSPNTPONTSGOICLVYTVPKDVVGOFAFFFTALNDVVEVDHGSOH 1440
 DB 1381 QYVGSDBVULSPNTPONTSGOICLVYTVPKDVVGOFAFFFTALNDVVEVDHGSOH 1440
 QY 1441 SRLLSLSGSHGTESLPLATSNQVLIKFSAGLAPAGFHFVQAVRTSATQCSSVPEP 1500
 DB 1441 SRLLSLSGSHGTESLPLATSNQVLIKFSAGLAPAGFHFVQAVRTSATQCSSVPEP 1500
 QY 1501 RYKGRILGSDFSVGAIVRFECSNGYALQGSPEIECLPVFGALAQNNVSAFPCVVPCCGNTL 1560
 DB 1501 RYKGRILGSDFSVGAIVRFECSNGYALQGSPEIECLPVFGALAQNNVSAFPCVVPCCGNTL 1560

QY 1561 ERGTLISPGFPEPYLNSLNCVWKIVPEBAGIQIOVSEFVTEBQNDLSLEVPDADNTYT 1620
 DB 1561 ERGTLISPGFPEPYLNSLNCVWKIVPEBAGIQIOVSEFVTEBQNDLSLEVPDADNTYT 1620
 QY 1621 MUGSFSGITVPALLNSTSNQVLIHAFYSDSVSAAGFHELYKTVGLSSCEBPAVNSGVKT 1680
 DB 1621 MUGSFSGITVPALLNSTSNQVLIHAFYSDSVSAAGFHELYKTVGLSSCEBPAVNSGVKT 1680
 QY 1681 GERLYVNDVVSFQCEBPGYALQGHANHSCMPGYTRANNYPPLICACCGTVEBMEGVILS 1740
 DB 1681 GERLYVNDVVSFQCEBPGYALQGHANHSCMPGYTRANNYPPLICACCGTVEBMEGVILS 1740
 QY 1741 PGFPNGYPSNMDCSMKIALPVGFGAIIOTLNFSTEBNHYIETIRNPTYETSRMAGFSS 1800
 DB 1741 PGFPNGYPSNMDCSMKIALPVGFGAIIOTLNFSTEBNHYIETIRNPTYETSRMAGFSS 1800
 QY 1801 ELPPSSLLSTSHETTVYFHSBDSGNRPGFLEYQAYELOECBPDEPPANGIVRAGYVNCQ 1860
 DB 1801 ELPPSSLLSTSHETTVYFHSBDSGNRPGFLEYQAYELOECBPDEPPANGIVRAGYVNCQ 1860
 QY 1861 SVTFECLPGYQLTGHVVLTCORGTNRNMDHPLPKCEVPGCGNITSSNGTYVSFSPSPYS 1920
 DB 1861 SVTFECLPGYQLTGHVVLTCORGTNRNMDHPLPKCEVPGCGNITSSNGTYVSFSPSPYS 1920
 QY 1921 SSODCVMLITVPIGHGVRLNLSLOTEPSGDFITINDGPOQTAPRLGVFTRSMAKKTVO 1980
 DB 1921 SSODCVMLITVPIGHGVRLNLSLOTEPSGDFITINDGPOQTAPRLGVFTRSMAKKTVO 1980
 QY 1981 SSNOVLLKPHRDAATGIFALIASAVPLTKCPPTLLPNAEVVTEBEEFNIGDIIVRYCL 2040
 DB 1981 SSNOVLLKPHRDAATGIFALIASAVPLTKCPPTLLPNAEVVTEBEEFNIGDIIVRYCL 2040
 QY 2041 PGFTLVGNELTKLCTGYLOFEGBPPICEVHCTNBLDSTVILISOSIPGYPQFOTC 2100
 DB 2041 PGFTLVGNELTKLCTGYLOFEGBPPICEVHCTNBLDSTVILISOSIPGYPQFOTC 2100
 QY 2101 SWLVREBPDNISLTYEYFSEKQYDEFEIPDPSQSOPLKALSGNYSAPLIVTSSNS 2160
 DB 2101 SWLVREBPDNISLTYEYFSEKQYDEFEIPDPSQSOPLKALSGNYSAPLIVTSSNS 2160
 QY 2161 VYLWMSDHAYNKGRKIRYSAPYCSLPRALHFIIGOTSTPGSIFHGACNAYRLVG 2220
 DB 2161 VYLWMSDHAYNKGRKIRYSAPYCSLPRALHFIIGOTSTPGSIFHGACNAYRLVG 2220
 QY 2221 HSMALICTRHPQGHLMSEAPILCOALSCEIPAPKNGMFGKTYTGTAXVYSCSSEGYL 2280
 DB 2221 HSMALICTRHPQGHLMSEAPILCOALSCEIPAPKNGMFGKTYTGTAXVYSCSSEGYL 2280
 QY 2281 QAGABATBECLEDTGLMNNRVPPQCVPTCCPDVSSISVEHGRRLIFETQYQFOAOLMLI 2340
 DB 2281 QAGABATBECLEDTGLMNNRVPPQCVPTCCPDVSSISVEHGRRLIFETQYQFOAOLMLI 2340
 QY 2341 CDPGYVYTGQVRLRCOANGKMSLGDSTPCRILISCEGLPIPNGBHRIIGTLVYGATAIFS 2400
 DB 2341 CDPGYVYTGQVRLRCOANGKMSLGDSTPCRILISCEGLPIPNGBHRIIGTLVYGATAIFS 2400
 QY 2401 CNGGTYLVGSRVBECMANGMSSBYRCLAGHGTBEPYVNGHINBENTSYRSVYQCN 2460
 DB 2401 CNGGTYLVGSRVBECMANGMSSBYRCLAGHGTBEPYVNGHINBENTSYRSVYQCN 2460
 QY 2461 AGFRLIGMSVRIQOQDHMSGKTPFCVPTICGHPGPNVGLTQGNQPNLNDVVKFCNPG 2520
 DB 2461 AGFRLIGMSVRIQOQDHMSGKTPFCVPTICGHPGPNVGLTQGNQPNLNDVVKFCNPG 2520
 QY 2521 YMAEGAARSGCLASGQMSDWLPTCRIINCTDPGHQENSVRQVHASPFRSFGTIVSYRC 2580
 DB 2521 YMAEGAARSGCLASGQMSDWLPTCRIINCTDPGHQENSVRQVHASPFRSFGTIVSYRC 2580
 QY 2581 NHGFTYLLGTPVLSQOQDGTWDRPPOCLVSCGHPSPSPHSONSGSYTYGAVVRYSCIG 2640
 DB 2581 NHGFTYLLGTPVLSQOQDGTWDRPPOCLVSCGHPSPSPHSONSGSYTYGAVVRYSCIG 2640
 QY 2641 KRTLVGNSTRMGLDGHMTGSLPHCSGTSVGVCGDPCGIPAHGIRLSDSPDPTVMEFSCE 2700

Db 2641 KRILVGNSTRMCGJGDHWTGSLPHCSGTSGVCCDPCGIPANGITLGSFDPGTMRSC 2700
Qy 2701 AGHYLRGSSERTCOANGSGSGOPEBCGYISCGNPQTSPNARVFSDLVFSSTIVYECRE 2760
Db 2701 AGHYLRGSSERTCOANGSGSGOPEBCGYISCGNPQTSPNARVFSDLVFSSTIVYECRE 2760
Qy 2761 GYVATGILSRHCSVNGTGTGSDPECLVINCSDPGIPANGILGNDPFXNTVYQCVPGY 2820
Db 2761 GYVATGILSRHCSVNGTGTGSDPECLVINCSDPGIPANGILGNDPFXNTVYQCVPGY 2820
Qy 2821 MMEHRSVLSCTYDRFTWNGTKPVCKALMCKPPLIENGKVGSDPFMGSSVTYACLEGY 2880
Db 2821 MMEHRSVLSCTYDRFTWNGTKPVCKALMCKPPLIENGKVGSDPFMGSSVTYACLEGY 2880
Qy 2881 QLSLPAVFTCGNSWNGELPQCFFPVCGDPRGVSRRDRGRSYSSVSFSGHPPLV 2940
Db 2881 QLSLPAVFTCGNSWNGELPQCFFPVCGDPRGVSRRDRGRSYSSVSFSGHPPLV 2940
Qy 2941 VGSRRFCQSDGTWGTGSDPTLTTCADPGVPGVPGIIONNSQGVGSTVLFRCKGY 3000
Db 2941 VGSRRFCQSDGTWGTGSDPTLTTCADPGVPGVPGIIONNSQGVGSTVLFRCKGY 3000
Qy 3001 LLOGSTRITCLPNTWGTGTPDDCVPHHCROPEFTPHANVAGALDLPMSGYTLITPARAS 3060
Db 3001 LLOGSTRITCLPNTWGTGTPDDCVPHHCROPEFTPHANVAGALDLPMSGYTLITPARAS 3060
Qy 3061 SRVAPSTAPARMAAGASRPSAMRSGVDPDPTLPSGSHSPK 3104
Db 3061 SRVAPSTAPARMAAGASRPSAMRSGVDPDPTLPSGSHSPK 3104

RESULT 3
ADH71138
ID ADH71138 standard; protein; 3104 AA.
XX
XX ADH71138;
XX
XX 25-MAR-2004 (first entry)
XX
XX Human protein of the invention NOVab SEQ ID NO:34.
XX
XX human; cytostatic; immunomodulator; neuroprotective; nootropic;
XX anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;
XX vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
XX obesity; diabetes; infectious disease; metabolic syndrome X;
XX dyslipidaemia.
XX
XX Homo sapiens.
XX
XX MO2003102155-A2.
XX
XX 11-DEC-2003.
XX
XX 03-JUN-2003; 2003MO-US017430.
XX
XX 03-JUN-2002; 2002US-0385120P.
XX 04-JUN-2002; 2002US-0385784P.
XX 05-JUN-2002; 2002US-0386041P.
XX 06-JUN-2002; 2002US-0386047P.
XX 06-JUN-2002; 2002US-0386376P.
XX 06-JUN-2002; 2002US-0386453P.
XX 06-JUN-2002; 2002US-038664P.
XX 06-JUN-2002; 2002US-0387016P.
XX 07-JUN-2002; 2002US-0386796P.
XX 07-JUN-2002; 2002US-0386816P.
XX 07-JUN-2002; 2002US-0386931P.
XX 07-JUN-2002; 2002US-0386942P.
XX 07-JUN-2002; 2002US-0386971P.
XX 07-JUN-2002; 2002US-0387262P.
XX 08-JUN-2002; 2002US-0387400P.
XX 10-JUN-2002; 2002US-0387460P.
XX 10-JUN-2002; 2002US-0387535P.

PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387696P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 17-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389844P.
PR 19-JUN-2002; 2002US-039006P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396766P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.
PR 09-AUG-2002; 2002US-0402256P.
PR 12-AUG-2002; 2002US-0402389P.
PR 12-AUG-2002; 2002US-0402786P.
PR 12-AUG-2002; 2002US-0402816P.
PR 12-AUG-2002; 2002US-0402821P.
PR 13-AUG-2002; 2002US-0402832P.
PR 13-AUG-2002; 2002US-0403448P.
PR 13-AUG-2002; 2002US-0403459P.
PR 13-AUG-2002; 2002US-0403531P.
PR 13-AUG-2002; 2002US-0403532P.
PR 13-AUG-2002; 2002US-0403563P.
PR 13-AUG-2002; 2002US-0406317P.
PR 15-AUG-2002; 2002US-0403617P.
PR 26-AUG-2002; 2002US-0406182P.
PR 26-AUG-2002; 2002US-0406355P.
PR 27-AUG-2002; 2002US-0406240P.
PR 12-SEP-2002; 2002US-0410084P.
PR 20-SEP-2002; 2002US-0412528P.
PR 23-SEP-2002; 2002US-0412731P.
PR 30-SEP-2002; 2002US-0414801P.
PR 30-SEP-2002; 2002US-0414839P.
PR 30-SEP-2002; 2002US-0414840P.
PR 30-SEP-2002; 2002US-0414954P.
PR 09-OCT-2002; 2002US-0417186P.
PR 09-OCT-2002; 2002US-0417406P.
PR 23-OCT-2002; 2002US-0420639P.
PR 28-OCT-2002; 2002US-0421156P.
PR 31-OCT-2002; 2002US-0422690P.
PR 01-NOV-2002; 2002US-0423130P.
PR 05-NOV-2002; 2002US-0423798P.
PR 05-NOV-2002; 2002US-0423798P.
PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.
XX
XX Alsebrook JP, Alvarez E, Anderson DM, Boldog FL, Casman SJ;
XX Caterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
XX Etemberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
XX Guev YV, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;
XX MacIsachlan T, Malayanar UM, Mezick AJ, Millet I, Mishra VS;
XX Padigaru M, Patturajan M, Pena CBA, Peyman JA, Raha D, Rastelli L;
XX Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
XX Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
XX Zhong H;
XX
XX WPI; 2004-081935/08.

DR N-PSDB; ADH71137.
 XX New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 XX Example 4; SEQ ID NO 34; 1880bp; English.
 CC The invention relates to a novel isolated polypeptide (NOVX). A
 CC polypeptide of the invention has cytosstatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antihypertensive activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOVX polypeptide of the invention.
 CC
 XX
 SQ Sequence 3104 AA;
 Query Match 100.0%; Score 16385; DB 8; Length 3104;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAGAPAPALLPCLSLSDCCASNQRHSGVSPSELVYKQIELKSRGVKMPKSDNSOKTS 60
 DB 1 MAGAPAPALLPCLSLSDCCASNQRHSGVSPSELVYKQIELKSRGVKMPKSDNSOKTS 60
 QY 61 VLTQVGVSQGHNNCPDPIPERGKRLGSDFLSSVQFTCEGYDLOGSKRITCMKVS DM 120
 DB 61 VLTQVGVSQGHNNCPDPIPERGKRLGSDFLSSVQFTCEGYDLOGSKRITCMKVS DM 120
 QY 121 FAAMSDHRYPCRAMCDNAH.RGPGIITTSNFP.IQYNNNAHCWITIALMPKVIKLA FE 180
 DB 121 FAAMSDHRYPCRAMCDNAH.RGPGIITTSNFP.IQYNNNAHCWITIALMPKVIKLA FE 180
 QY 181 EFDLERGYDTLTVGDGQDQDKTVLYMSQNA.CSDSPHTPGSRI.PESMSGDIMROKVT V 240
 DB 181 EFDLERGYDTLTVGDGQDQDKTVLYMSQNA.CSDSPHTPGSRI.PESMSGDIMROKVT V 240
 QY 241 EICRDISSDARSQSVKSPKTSNAVELVAPGTEIEGSCGDPGIPAYGRREGSRFTHGD 300
 DB 241 EICRDISSDARSQSVKSPKTSNAVELVAPGTEIEGSCGDPGIPAYGRREGSRFTHGD 300
 QY 301 TLKFECPAPFELVQKAI.TQKNNQMSAKKPGCVFSCFENFTSPSGVVLSPNTPEDYGNH 360
 DB 301 TLKFECPAPFELVQKAI.TQKNNQMSAKKPGCVFSCFENFTSPSGVVLSPNTPEDYGNH 360
 QY 361 LHCWMLILAPRESRIHAFNDIYEPQDFLVIKOGATAEPVLTGTFSGNQLSSITSSG 420
 DB 361 LHCWMLILAPRESRIHAFNDIYEPQDFLVIKOGATAEPVLTGTFSGNQLSSITSSG 420
 QY 421 HVARLEFQDHSHTGKRGFNITFTFRNNECPDPGPVNVGKRFQDLSGLSSISPLCDGEGF 480
 DB 421 HVARLEFQDHSHTGKRGFNITFTFRNNECPDPGPVNVGKRFQDLSGLSSISPLCDGEGF 480
 QY 481 LGTQGSSTITCVLKEGVSVMNSAVLRCEAPCGHLTSPSGTILSPGPGYKALSCAWV 540
 DB 481 LGTQGSSTITCVLKEGVSVMNSAVLRCEAPCGHLTSPSGTILSPGPGYKALSCAWV 540
 QY 541 IEAOPGPKITFDRFTEVNVDTLVRDGRITSAPLIGYHGTQVQFPISTSNLYL 600
 DB 541 IEAOPGPKITFDRFTEVNVDTLVRDGRITSAPLIGYHGTQVQFPISTSNLYL 600
 QY 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI.PVNGQRHND.FYVALVTSCDSGYTLSD 660
 DB 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGI.PVNGQRHND.FYVALVTSCDSGYTLSD 660

QY 661 GEPLCEBPNOFWRALPSCALCGFLOQSSGTLSPGPPDPYNNLNTCTIETSHGK 720
 DB 661 GEPLCEBPNOFWRALPSCALCGFLOQSSGTLSPGPPDPYNNLNTCTIETSHGK 720
 QY 721 VFFTFHTFHLSESHDVLITENGSTPQPLRLTGSRLPAPISAGLVNFTQVRFISDFS 780
 DB 721 VFFTFHTFHLSESHDVLITENGSTPQPLRLTGSRLPAPISAGLVNFTQVRFISDFS 780
 QY 781 MSYEGNITTSYEDLEPCEBPVAYSIRKGLQFGVDLTTCSPFGYRLGRTARITCLG 840
 DB 781 MSYEGNITTSYEDLEPCEBPVAYSIRKGLQFGVDLTTCSPFGYRLGRTARITCLG 840
 QY 841 GRRRLMSSPLPRCAVACGNSVTGQTLSPNPNVNNHETIYSIQTPQKGIOUKAR 900
 DB 841 GRRRLMSSPLPRCAVACGNSVTGQTLSPNPNVNNHETIYSIQTPQKGIOUKAR 900
 QY 901 AFELSGDVLKYVDGNNSARLLGVFSHSEMMGVLTNTSSSLMDPITDAENTSKGFE 960
 DB 901 AFELSGDVLKYVDGNNSARLLGVFSHSEMMGVLTNTSSSLMDPITDAENTSKGFE 960
 QY 961 HFSFELIKCEDPTKFGYKVDDEGHFAGSSVSFSCDPGYSLRGSBELCLSGERTWD 1020
 DB 961 HFSFELIKCEDPTKFGYKVDDEGHFAGSSVSFSCDPGYSLRGSBELCLSGERTWD 1020
 QY 1021 RPLPTCAVBCGTVRGEVSGVLSPGYAPAYENHNCIMTIEBACCTIGLHFLVPTDE 1080
 DB 1021 RPLPTCAVBCGTVRGEVSGVLSPGYAPAYENHNCIMTIEBACCTIGLHFLVPTDE 1080
 QY 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDLHSTFNSVVLQFTDFTSKQGFALIOFSVST 1140
 DB 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDLHSTFNSVVLQFTDFTSKQGFALIOFSVST 1140
 QY 1141 ATSCNDPGLPONGSRGDSWEAGDSTVPOCDPGYALQGSALISCVKIEENFMQPSPTC 1200
 DB 1141 ATSCNDPGLPONGSRGDSWEAGDSTVPOCDPGYALQGSALISCVKIEENFMQPSPTC 1200
 QY 1201 IAPCGDILGPGSVLILSPNYPPEYPPGKEDMKVTUSPDVIALVNIINFLFEGYFLHI 1260
 DB 1201 IAPCGDILGPGSVLILSPNYPPEYPPGKEDMKVTUSPDVIALVNIINFLFEGYFLHI 1260
 QY 1261 YDGRDLSPLIGSFYSGQLPGR.IESSNSL.FLAFRSADASYNAGFYIDYENPRESGFD 1320
 DB 1261 YDGRDLSPLIGSFYSGQLPGR.IESSNSL.FLAFRSADASYNAGFYIDYENPRESGFD 1320
 QY 1321 GSINKGTRVGS.DIKLSSVTVYCHGGYEVGISTLSCLIGPDGKPYWNNRPVCTAPCGG 1380
 DB 1321 GSINKGTRVGS.DIKLSSVTVYCHGGYEVGISTLSCLIGPDGKPYWNNRPVCTAPCGG 1380
 QY 1381 QYVSGDVVLS.PYVPONTYSGOI.CLYFVTVPKDYVVGQAFPHTLANDVVEVHDGSHQ 1440
 DB 1381 QYVSGDVVLS.PYVPONTYSGOI.CLYFVTVPKDYVVGQAFPHTLANDVVEVHDGSHQ 1440
 QY 1441 SRLSSLSGSHGESLPLATSNQVLI.KFSAKGLAPARGHFVYQAVPRTSATOCSSVPER 1500
 DB 1441 SRLSSLSGSHGESLPLATSNQVLI.KFSAKGLAPARGHFVYQAVPRTSATOCSSVPER 1500
 QY 1501 RYKRLGSDPFSVGAIVRPECNSGYALQGSPEIECLVPGLAQMNVSAPTCVPPCGANLT 1560
 DB 1501 RYKRLGSDPFSVGAIVRPECNSGYALQGSPEIECLVPGLAQMNVSAPTCVPPCGANLT 1560
 QY 1561 ERRTGTLSPGPEPYNLNSLNCWKIVPBRAGIQIVVSVFVIRQNNDSLEVFAGANTVY 1620
 DB 1561 ERRTGTLSPGPEPYNLNSLNCWKIVPBRAGIQIVVSVFVIRQNNDSLEVFAGANTVY 1620
 QY 1621 MLGSFSGTTPALNLTNSQLYLH.FYSDSLVSNAAGFHEVTKVTGLSSCEBPAPSNAGVT 1680
 DB 1621 MLGSFSGTTPALNLTNSQLYLH.FYSDSLVSNAAGFHEVTKVTGLSSCEBPAPSNAGVT 1680
 QY 1681 GERLYLVNDVVS.FOCBEGYALQGHARI.SCMPTVRRNNYPPPLCTAOCGGTVEBMEGVILS 1740
 DB 1681 GERLYLVNDVVS.FOCBEGYALQGHARI.SCMPTVRRNNYPPPLCTAOCGGTVEBMEGVILS 1740


```

QY 1741 PGFGPNYNSMDCWKIALPVFGAHIQFLNFSTBPNHDYIEIRNGEPYTSRMKGRSGS 1800
DB 1741 PGFGPNYNSMDCWKIALPVFGAHIQFLNFSTBPNHDYIEIRNGEPYTSRMKGRSGS 1800
QY 1801 ELPSLSLSTSHETTYVPHSDHSQNRPGFKLEYAYELQECBPDPFPFANGIVRGAGYVNG 1860
DB 1801 ELPSLSLSTSHETTYVPHSDHSQNRPGFKLEYAYELQECBPDPFPFANGIVRGAGYVNG 1860
QY 1861 SVTEECIPGVOLJTHGHPVLTQCHGTRNRMNDHLPCEVPCCGNITSSNGITYSFGPSPYS 1920
DB 1861 SVTEECIPGVOLJTHGHPVLTQCHGTRNRMNDHLPCEVPCCGNITSSNGITYSFGPSPYS 1920
QY 1921 SSODCWLITVPIGHGVRNLNLSLQTEPSGDFITIMDGPOTAARLVEFTRSMAKKTVOS 1980
DB 1921 SSODCWLITVPIGHGVRNLNLSLQTEPSGDFITIMDGPOTAARLVEFTRSMAKKTVOS 1980
QY 1981 SSNOVLLKFRHDAATGGTIPAIASFAYPLTKCPPTIIPNAEVTENEENIGDIVRYCL 2040
DB 1981 SSNOVLLKFRHDAATGGTIPAIASFAYPLTKCPPTIIPNAEVTENEENIGDIVRYCL 2040
QY 2041 PGFTLVGNELITCKLGTYLQEPGPPICEVHCPTNELITDSTGVILISQSPGSPYPOFQC 2100
DB 2041 PGFTLVGNELITCKLGTYLQEPGPPICEVHCPTNELITDSTGVILISQSPGSPYPOFQC 2100
QY 2101 SWLVREPDYNISLTVYFLSEKQYDFEIPDGPSSQPLKALSGNYSAPLIVTSSNS 2160
DB 2101 SWLVREPDYNISLTVYFLSEKQYDFEIPDGPSSQPLKALSGNYSAPLIVTSSNS 2160
QY 2161 VYLWMSDHAYNRGKFIKIRYAPYCSLPAPLHGFLIGOTSTOPGSIHFGCNAGRYLVG 2220
DB 2161 VYLWMSDHAYNRGKFIKIRYAPYCSLPAPLHGFLIGOTSTOPGSIHFGCNAGRYLVG 2220
QY 2221 HSNALCTRHPOGYHLMSEAIPLCOALSCGLPEAPKNGMVGKXYTGTVKAVYSCSEGYH 2280
DB 2221 HSNALCTRHPOGYHLMSEAIPLCOALSCGLPEAPKNGMVGKXYTGTVKAVYSCSEGYH 2280
QY 2281 QAQAENATAECLDTGLMSNNRVPPQCVPTCPDVSSISVEHGRNLLIPEYOQFOAQLMI 2340
DB 2281 QAQAENATAECLDTGLMSNNRVPPQCVPTCPDVSSISVEHGRNLLIPEYOQFOAQLMI 2340
QY 2341 CDPGYVYTGQRYVRCOANGKMSLGDSTPTCRITISCEBLPPNGHRICTLVYGAATAS 2400
DB 2341 CDPGYVYTGQRYVRCOANGKMSLGDSTPTCRITISCEBLPPNGHRICTLVYGAATAS 2400
QY 2401 CNGSYTLVGSRYVRCOANGKMSLGDSTPTCRITISCEBLPPNGHRICTLVYGAATAS 2460
DB 2401 CNGSYTLVGSRYVRCOANGKMSLGDSTPTCRITISCEBLPPNGHRICTLVYGAATAS 2460
QY 2461 AGFRLIGMSVRLICQDHHMSGKTPFCVPTTCGHPNVNGLTOGNOFNINDVVKFYCNFG 2520
DB 2461 AGFRLIGMSVRLICQDHHMSGKTPFCVPTTCGHPNVNGLTOGNOFNINDVVKFYCNFG 2520
QY 2521 YMAEGARSOCLASGQMDMLPTCRIINCTDPGHQNSROYVHASGPHRFSGTIVSYNC 2580
DB 2521 YMAEGARSOCLASGQMDMLPTCRIINCTDPGHQNSROYVHASGPHRFSGTIVSYNC 2580
QY 2581 NHGFYLLGTPVYLSQGDGTWDRPRPQCLLVSCGHPSPHSQMSGDSYTVGAVVYXSCIG 2640
DB 2581 NHGFYLLGTPVYLSQGDGTWDRPRPQCLLVSCGHPSPHSQMSGDSYTVGAVVYXSCIG 2640
QY 2641 KRTLIVGNSTRMGLDGHMTGSLPHGSGTVGCGDGTGAHGRILGLDSDPCTVWRFSCE 2700
DB 2641 KRTLIVGNSTRMGLDGHMTGSLPHGSGTVGCGDGTGAHGRILGLDSDPCTVWRFSCE 2700
QY 2701 AGHVLARSSERTQANGSWSGQPECGVISCNPGTPSNARVVFSDGLVFSSTIYECRE 2760
DB 2701 AGHVLARSSERTQANGSWSGQPECGVISCNPGTPSNARVVFSDGLVFSSTIYECRE 2760
QY 2761 GYATATLLRHSVNGTWTGSDPECLVINGDPGIPANGRLGDNFRINKTVTYOCVPGY 2820
DB 2761 GYATATLLRHSVNGTWTGSDPECLVINGDPGIPANGRLGDNFRINKTVTYOCVPGY 2820
QY 2821 MMESHRSVLSCTKRTVNGTKPVCKALMCKPPLIPNGKVVSGSPMMGSSVTVACLEGY 2880

```

```

DB 2821 MMESHRSVLSCTKRTVNGTKPVCKALMCKPPLIPNGKVVSGSPMMGSSVTVACLEGY 2880
QY 2881 QLSLPAVFCENGSWGTGELPQCFPVFCGDPVSPRGREDGFSYRSSVSCHPPLVL 2940
DB 2881 QLSLPAVFCENGSWGTGELPQCFPVFCGDPVSPRGREDGFSYRSSVSCHPPLVL 2940
QY 2941 VSPRRFCOSDGTWSGTQPSCIDPILTTCADBGVPOFGIUNNSQGYQVGSYVLPFCQKGY 3000
DB 2941 VSPRRFCOSDGTWSGTQPSCIDPILTTCADBGVPOFGIUNNSQGYQVGSYVLPFCQKGY 3000
QY 3001 LLOGSTTRTCLPNLWMSGTPDCVPHHCRQEPPTHANVAGLIDPSMGYTLITPARASP 3060
DB 3001 LLOGSTTRTCLPNLWMSGTPDCVPHHCRQEPPTHANVAGLIDPSMGYTLITPARASP 3060
QY 3061 SRVASTAPARRMAGASRPSAMRSRPSGVDPSTLPGSHRSFKP 3104
DB 3061 SRVASTAPARRMAGASRPSAMRSRPSGVDPSTLPGSHRSFKP 3104

RESULT 4
ADH71168
ID ADH71168 standard; protein; 3104 AA.
XX
AC ADH71168;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human protein of the invention NOV4q SEQ ID NO:64.
XX
KW human; cytostatic; immunomodulator; neuroprotective; nootropic;
KW anorectic; antidiabetic; antimicrobial; antilipæmic; gene therapy;
KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
KW obesity; diabetes; infectious disease; metabolic syndrome X;
KW dyslipidaemia.
XX
OS Homo sapiens.
XX
PN WO2003102155-A2.
XX
PD 11-DEC-2003.
XX
PF 03-JUN-2003; 2003WO-US017430.
XX
PR 03-JUN-2002; 2002US-0385710P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 06-JUN-2002; 2002US-0386376P.
PR 07-JUN-2002; 2002US-0386453P.
PR 08-JUN-2002; 2002US-0386864P.
PR 09-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-038796P.
PR 07-JUN-2002; 2002US-038816P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0386942P.
PR 07-JUN-2002; 2002US-0386971P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-0387625P.
PR 10-JUN-2002; 2002US-0387400P.
PR 10-JUN-2002; 2002US-0387553P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387686P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.

```

PR 12-JUN-2002; 2002US-0386096P.
 PR 13-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389118P.
 PR 14-JUN-2002; 2002US-0389120P.
 PR 14-JUN-2002; 2002US-0389144P.
 PR 14-JUN-2002; 2002US-0389146P.
 PR 17-JUN-2002; 2002US-0389729P.
 PR 17-JUN-2002; 2002US-0389742P.
 PR 18-JUN-2002; 2002US-0389884P.
 PR 19-JUN-2002; 2002US-0390006P.
 PR 19-JUN-2002; 2002US-0390209P.
 PR 21-JUN-2002; 2002US-0390763P.
 PR 17-JUL-2002; 2002US-0396706P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402156P.
 PR 09-AUG-2002; 2002US-0402256P.
 PR 09-AUG-2002; 2002US-0402389P.
 PR 12-AUG-2002; 2002US-0402786P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 12-AUG-2002; 2002US-0402832P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403531P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 13-AUG-2002; 2002US-0403577P.
 PR 15-AUG-2002; 2002US-0403617P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-0412731P.
 PR 30-SEP-2002; 2002US-0414801P.
 PR 30-SEP-2002; 2002US-0414839P.
 PR 30-SEP-2002; 2002US-0414840P.
 PR 30-SEP-2002; 2002US-0414954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Alsbrook JP, Alvarez E, Anderson DW, Boldog FL, Caeman SJ,
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,
 PI Etienne S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X,
 PI Guev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,
 PI McLachlan T, Malvanter UM, Meick AJ, Millet I, Mishra VS,
 PI Padgett M, Patutajan M, Pena CE, Peyman JA, Raha D, Rastelli L,
 PI Rieger DK, Rotenberg ME, Scioe P, Shenoy SG, Shimkets RA,
 PI Smithson G, Spletter KA, Stone DJ, Vernet CM, Voss EZ, Zhong M,
 PI Zhong H;
 XX
 DR WPI: 2004-081935/08.
 DR N-PSDB: ADH71167.
 XX
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 XX Example 4, SEQ ID NO 64; 1880bp; English.
 CC
 CC The invention relates to a novel isolated polypeptide (NOVX). A
 CC polypeptide of the invention has cytosstatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antilipemic activity, and may have a use in gene therapy, and as a

CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOVX polypeptide of the invention.
 XX
 SQ Sequence 3104 AA;
 Query Match 100.0%; Score 16978; DB 8; Length 3104;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3103; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MAGAPPALLPCLSLSDCCASNORHVGVPSELVKKQIELKSRGVKLMPSKDNISQKTS 60
 DB 1 MAGAPPALLPCLSLSDCCASNORHVGVPSELVKKQIELKSRGVKLMPSKDNISQKTS 60
 QY VLTQVGVSGHHNCPDGPGERGRGLSDPRLGSSVQFTCNBEGYDLQSGKRITCMKVS DM 120
 DB VLTQVGVSGHHNCPDGPGERGRGLSDPRLGSSVQFTCNBEGYDLQSGKRITCMKVS DM 120
 QY 61 VLTQVGVSGHHNCPDGPGERGRGLSDPRLGSSVQFTCNBEGYDLQSGKRITCMKVS DM 120
 DB 61 VLTQVGVSGHHNCPDGPGERGRGLSDPRLGSSVQFTCNBEGYDLQSGKRITCMKVS DM 120
 QY 121 FAAMSDHPRVCRAMCMCAHRLRSGGIIITSNFPIQDNNNAHCWIIITALLMPSKYIKLAF 180
 DB 121 FAAMSDHPRVCRAMCMCAHRLRSGGIIITSNFPIQDNNNAHCWIIITALLMPSKYIKLAF 180
 QY 181 EFDLERGYDTLTVGGDQDDOKTVLMSGNACSDSPHTGSRIPESMSGDIWROKTVL 240
 DB 181 EFDLERGYDTLTVGGDQDDOKTVLMSGNACSDSPHTGSRIPESMSGDIWROKTVL 240
 QY 181 EFDLERGYDTLTVGGDQDDOKTVLMSGNACSDSPHTGSRIPESMSGDIWROKTVL 240
 DB 181 EFDLERGYDTLTVGGDQDDOKTVLMSGNACSDSPHTGSRIPESMSGDIWROKTVL 240
 QY 241 EICRDISSDARSQSVKSPKTSNAVELVAPTEIEGSGCDPGIIPAYGRREGSRFHGD 300
 DB 241 EICRDISSDARSQSVKSPKTSNAVELVAPTEIEGSGCDPGIIPAYGRREGSRFHGD 300
 QY 241 EICRDISSDARSQSVKSPKTSNAVELVAPTEIEGSGCDPGIIPAYGRREGSRFHGD 300
 DB 241 EICRDISSDARSQSVKSPKTSNAVELVAPTEIEGSGCDPGIIPAYGRREGSRFHGD 300
 QY 301 TLKFECPAPFELVQKAITCKKNNQSAKPGCVFSCFFNFPTSPSGVLLPNYPEDYGNH 360
 DB 301 TLKFECPAPFELVQKAITCKKNNQSAKPGCVFSCFFNFPTSPSGVLLPNYPEDYGNH 360
 QY 301 TLKFECPAPFELVQKAITCKKNNQSAKPGCVFSCFFNFPTSPSGVLLPNYPEDYGNH 360
 DB 301 TLKFECPAPFELVQKAITCKKNNQSAKPGCVFSCFFNFPTSPSGVLLPNYPEDYGNH 360
 QY 361 LHCWMLILARPESRIHLAFNDIVPEQDFLVIKDGTAAAPVLGTSGNQOLPSSITSSG 420
 DB 361 LHCWMLILARPESRIHLAFNDIVPEQDFLVIKDGTAAAPVLGTSGNQOLPSSITSSG 420
 QY 421 HVARLEFQDHTSGTKRGFNITFTFRNECPDGPVPNGKRFQDLSQLGSSISFLCDEGF 480
 DB 421 HVARLEFQDHTSGTKRGFNITFTFRNECPDGPVPNGKRFQDLSQLGSSISFLCDEGF 480
 QY 481 LGTQSGRTITCVLKEGSVVNSAVLCEAPCGHLISPSGIIISPCMPGFKALSCAMV 540
 DB 481 LGTQSGRTITCVLKEGSVVNSAVLCEAPCGHLISPSGIIISPCMPGFKALSCAMV 540
 QY 541 IBAQPGYKITDRPFTEYVNDILEVRDRTYAPLIGYVHGVOQFLLISTSNVLYL 600
 DB 541 IBAQPGYKITDRPFTEYVNDILEVRDRTYAPLIGYVHGVOQFLLISTSNVLYL 600
 QY 601 FSTDKSHSDIGFQRLYETITLQSDHCLDPGIPVNGQRHNDFFVYALVTSCSGYTLSD 660
 DB 601 FSTDKSHSDIGFQRLYETITLQSDHCLDPGIPVNGQRHNDFFVYALVTSCSGYTLSD 660
 QY 661 GEPLECEPNTQMSBALPSCBALGGFQSSGGIITLSPGFDFPNNINCTWIIITSHGKG 720
 DB 661 GEPLECEPNTQMSBALPSCBALGGFQSSGGIITLSPGFDFPNNINCTWIIITSHGKG 720
 QY 721 VPFTHFTHESGHDYLLITENGSPFQPLQLGSRIPAPISAGLXNFAQVRFIDFS 780
 DB 721 VPFTHFTHESGHDYLLITENGSPFQPLQLGSRIPAPISAGLXNFAQVRFIDFS 780
 QY 781 MSYEGFNITSEYDLBCEBEPVAYSIRKGLQGVGDTLTFSCFPGRLEGARITCLG 840
 DB 781 MSYEGFNITSEYDLBCEBEPVAYSIRKGLQGVGDTLTFSCFPGRLEGARITCLG 840

QY 841 GRRILMSPLPRCAVECGNSVYTGOTLLSPNFPYNNNNHCTYSIOTOPGKIQOKAR 900
DB 841 GRRILMSPLPRCAVECGNSVYTGOTLLSPNFPYNNNNHCTYSIOTOPGKIQOKAR 900
QY 901 AFELSEGDLKYYDGNNSARLLGVFSHSEMMGVTLNSTSSSLMLDFTDAENTSKGPEL 960
DB 901 AFELSEGDLKYYDGNNSARLLGVFSHSEMMGVTLNSTSSSLMLDFTDAENTSKGPEL 960
QY 961 HPSSELLKCHDPGPKFGYKVDGHRAGSSVSFCDPGYSILGSEELLCSGERRTWD 1020
DB 961 HPSSELLKCHDPGPKFGYKVDGHRAGSSVSFCDPGYSILGSEELLCSGERRTWD 1020
QY 1021 RPLPTCAVECGGTARGEVSGQVLSFGYAPAEHNLNCITWEAAGCTTGLHFLVPTEE 1080
DB 1021 RPLPTCAVECGGTARGEVSGQVLSFGYAPAEHNLNCITWEAAGCTTGLHFLVPTEE 1080
QY 1081 VHDVLRITWDGVEGVLKELISGPALPKDLHSTNSVVLQFSTDFTSKGFAIQFSVST 1140
DB 1081 VHDVLRITWDGVEGVLKELISGPALPKDLHSTNSVVLQFSTDFTSKGFAIQFSVST 1140
QY 1141 ATSCNDPGIPONGSRSGDSWEAGDSTVFCODPGYALQGSABISGVKLENNFPMQSPPTC 1200
DB 1141 ATSCNDPGIPONGSRSGDSWEAGDSTVFCODPGYALQGSABISGVKLENNFPMQSPPTC 1200
QY 1201 IAPCGGDLTGSPSVILSPNYPEPYPGKECMDKVTSPDYIALVENIENLEPGYDPLH 1260
DB 1201 IAPCGGDLTGSPSVILSPNYPEPYPGKECMDKVTSPDYIALVENIENLEPGYDPLH 1260
QY 1261 YDGRDLSPLIGSVYSGOLPGRITSSNSLFLARSDASVSNAGFYVDYENRESCFDP 1320
DB 1261 YDGRDLSPLIGSVYSGOLPGRITSSNSLFLARSDASVSNAGFYVDYENRESCFDP 1320
QY 1321 GSINGRVGSDDLKGSSTVYCHGVEVEGTSTLSCITLGPDPKPYNNRPRVCTAPCG 1380
DB 1321 GSINGRVGSDDLKGSSTVYCHGVEVEGTSTLSCITLGPDPKPYNNRPRVCTAPCG 1380
QY 1381 QYVSDGVLSNPNPONTGQICLYVTVPKDYVFGQAFPHITLANDVEVHDHSHOH 1440
DB 1381 QYVSDGVLSNPNPONTGQICLYVTVPKDYVFGQAFPHITLANDVEVHDHSHOH 1440
QY 1441 SRIILSSGSHTEGSLPLATSNOLIKFSAKGLAPARGHFNVOANPRISATOCSSVBP 1500
DB 1441 SRIILSSGSHTEGSLPLATSNOLIKFSAKGLAPARGHFNVOANPRISATOCSSVBP 1500
QY 1501 RYKRLGSDPSVGAIVAFECNSGYALQGSPEIECLVPGLAQMNVSATCTVPCGGLT 1560
DB 1501 RYKRLGSDPSVGAIVAFECNSGYALQGSPEIECLVPGLAQMNVSATCTVPCGGLT 1560
QY 1561 ERRGTLSPGPEPYNLSLNCWKIYVBPAGAGIQIOVSVFTEQNMDSLEVFADANTVT 1620
DB 1561 ERRGTLSPGPEPYNLSLNCWKIYVBPAGAGIQIOVSVFTEQNMDSLEVFADANTVT 1620
QY 1621 MGSFSGTTPALNSTSNOLYHAFSDISVSAAGHLIKYKIVGSSCCEBPVPSNGVT 1680
DB 1621 MGSFSGTTPALNSTSNOLYHAFSDISVSAAGHLIKYKIVGSSCCEBPVPSNGVT 1680
QY 1681 GERIYLVNDVVSFCCEGYALQGHAIISCMPTVRANVYPPPLCIAQCGGTVEEMGVLLS 1740
DB 1681 GERIYLVNDVVSFCCEGYALQGHAIISCMPTVRANVYPPPLCIAQCGGTVEEMGVLLS 1740
QY 1741 PGPGNYPSNMDCSMKIALPVGGAHIOFLNSTENHDIYIIRNGPYETSMMRFGS 1800
DB 1741 PGPGNYPSNMDCSMKIALPVGGAHIOFLNSTENHDIYIIRNGPYETSMMRFGS 1800
QY 1801 ELPPSLISTSHETTYVFSHDSONRPGFLYQAEVLOECPPPEFANGIYVAGAYNGO 1860
DB 1801 ELPPSLISTSHETTYVFSHDSONRPGFLYQAEVLOECPPPEFANGIYVAGAYNGO 1860
QY 1861 SVTFECLPGYOLIGHVVLTCOHGTRNMDHPLPKCEVPCCGNITSSNGTIVSPGSPYS 1920
DB 1861 SVTFECLPGYOLIGHVVLTCOHGTRNMDHPLPKCEVPCCGNITSSNGTIVSPGSPYS 1920

QY 1921 SSODCWLLITVPIGHVRLNLSLQTEPSGDEITLWDGPOOTAPRLGVFTRSMAKTVOS 1980
DB 1921 SSODCWLLITVPIGHVRLNLSLQTEPSGDEITLWDGPOOTAPRLGVFTRSMAKTVOS 1980
QY 1981 SSNOVLKKEHRAATGCIFAIPASVAPLTKCPPTLLPNAEVTEENEFNIDIVRYRCL 2040
DB 1981 SSNOVLKKEHRAATGCIFAIPASVAPLTKCPPTLLPNAEVTEENEFNIDIVRYRCL 2040
QY 2041 PGFTLVGNELITCKLTGYLQFEGPPICEVHCPTMELLTDSTGYLLSOSYPSYQFQTC 2100
DB 2041 PGFTLVGNELITCKLTGYLQFEGPPICEVHCPTMELLTDSTGYLLSOSYPSYQFQTC 2100
QY 2101 SMLVREBPYNSLVEYFLSEKQYDEBEIFDPGSGOSPLKALSGNSAPLIVTSSNS 2160
DB 2101 SMLVREBPYNSLVEYFLSEKQYDEBEIFDPGSGOSPLKALSGNSAPLIVTSSNS 2160
QY 2161 VILRMSDHAHYRKFPKIRYSAPYCSLPRAPLHGTLLOTSOPGSGIHFPCNAAYRLVG 2220
DB 2161 VILRMSDHAHYRKFPKIRYSAPYCSLPRAPLHGTLLOTSOPGSGIHFPCNAAYRLVG 2220
QY 2221 HSMALCTHPOGYHLMSEBAPLQALSCGLPEAPKQNVFGKELYVGTKAVYCSGTYL 2280
DB 2221 HSMALCTHPOGYHLMSEBAPLQALSCGLPEAPKQNVFGKELYVGTKAVYCSGTYL 2280
QY 2281 QAGABATBCLDTGMSNRNVPQCVPTCDVSSISYEHGRMLIFETQYQFOQLMLI 2340
DB 2281 QAGABATBCLDTGMSNRNVPQCVPTCDVSSISYEHGRMLIFETQYQFOQLMLI 2340
QY 2341 CDPGYTYTGORYIRCOANGKSLGDSPTCRLISGELPIPNNGRICTLTVYGATAFS 2400
DB 2341 CDPGYTYTGORYIRCOANGKSLGDSPTCRLISGELPIPNNGRICTLTVYGATAFS 2400
QY 2401 CNSGTYLVGSVREEMANGLMSGSEVRLACHCCTPEPIYNGHINENYSYRSGVYQCN 2460
DB 2401 CNSGTYLVGSVREEMANGLMSGSEVRLACHCCTPEPIYNGHINENYSYRSGVYQCN 2460
QY 2461 AGFRLIGMSVRIQODHMSGKTPFCVPTITGHPGNPNGLTQONONLNDVNFVCNPG 2520
DB 2461 AGFRLIGMSVRIQODHMSGKTPFCVPTITGHPGNPNGLTQONONLNDVNFVCNPG 2520
QY 2521 YMAEGAARSQCLASGQMSDMLPTCRIINCTDPGHQENSVMQVHAGSHRHSFGITVYRC 2580
DB 2521 YMAEGAARSQCLASGQMSDMLPTCRIINCTDPGHQENSVMQVHAGSHRHSFGITVYRC 2580
QY 2581 NHGFYLLGTPLYSCOGSGTMDRPRPOCLLYSCGHPSPPHSOMSGDSTYGAIVRYCIG 2640
DB 2581 NHGFYLLGTPLYSCOGSGTMDRPRPOCLLYSCGHPSPPHSOMSGDSTYGAIVRYCIG 2640
QY 2641 KRLVGNSTRMCGLDGHWGSLPHCSGTSVGVCDPGI PAHGIKRLGDSFPGTYMRFSCE 2700
DB 2641 KRLVGNSTRMCGLDGHWGSLPHCSGTSVGVCDPGI PAHGIKRLGDSFPGTYMRFSCE 2700
QY 2701 AGHVLBSSERTCOANGSMGSOPEGCVISCGNPGTPSNARVYFSDGLVFSSIVYCRE 2760
DB 2701 AGHVLBSSERTCOANGSMGSOPEGCVISCGNPGTPSNARVYFSDGLVFSSIVYCRE 2760
QY 2761 GYVATGLSHRCSYNGMTGSDPECLVINGDGIIPANGRLANDPNPNKNTVYQCCPGY 2820
DB 2761 GYVATGLSHRCSYNGMTGSDPECLVINGDGIIPANGRLANDPNPNKNTVYQCCPGY 2820
QY 2821 MESHRYVSLCTKDRMTNGTKVCAIMCKPPLIPNGVYVSDMFMGSSVYALCLEG 2880
DB 2821 MESHRYVSLCTKDRMTNGTKVCAIMCKPPLIPNGVYVSDMFMGSSVYALCLEG 2880
QY 2881 QLSLPAFTCEBAGSWTGLPQCFPVFCBPGVPSGRREDRFFSYRSSFSCHPPLVY 2940
DB 2881 QLSLPAFTCEBAGSWTGLPQCFPVFCBPGVPSGRREDRFFSYRSSFSCHPPLVY 2940
QY 2941 VGSPPRFQCSGDSGTMSGTOSCIDPTLTTCCADPVVPFGI ONNSOGYOVSTYLFRCOKY 3000
DB 2941 VGSPPRFQCSGDSGTMSGTOSCIDPTLTTCCADPVVPFGI ONNSOGYOVSTYLFRCOKY 3000
QY 3001 LLOGSTTRCTLPLNLWSTGTPDCVPHCRQPETPTHANVALDLPSMGYTLITPARASP 3060

Db	3001	LLGGSTRRLPLNLTSGTTPDCVPHHCROPETPTANWGLDLPMMGTLITPARRSP	3060
Qy	3061	SRVAPSTAPARRMAAGASRPASMRGSPVDPSTLPGSHRSPRP	3104
Db	3061	SRVAPSTAPARRMAAGASRPASMRGSPVDPSTLPGSHRSPRP	3104
RESULT 5			
ADH71136			
ID	ADH71136	standard; protein; 3546 AA.	
AC	ADH71136;		
XX			
XX	25-MAR-2004	(first entry)	
DE		Human protein of the invention NOVA SEQ ID NO:32.	
XX			
XX		human; cytostatic; immunomodulator; neuroprotective; nootropic;	
XX		anorectic; antidiabetic; antimicrobial; antilipemic; gene therapy;	
XX		vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;	
XX		obesity; diabetes; infectious disease; metabolic syndrome X;	
XX		dyslipidaemia.	
OS	Homo sapiens.		
XX	WO2003102155-A2.		
XX			
PD	11-DEC-2003.		
XX			
XX	03-JUN-2003;	2003WO-US017430.	
XX			
PR	03-JUN-2002;	2002US-0385120P.	
PR	04-JUN-2002;	2002US-0385784P.	
PR	05-JUN-2002;	2002US-0386041P.	
PR	05-JUN-2002;	2002US-0386047P.	
PR	06-JUN-2002;	2002US-0386376P.	
PR	06-JUN-2002;	2002US-0386453P.	
PR	06-JUN-2002;	2002US-0386864P.	
PR	06-JUN-2002;	2002US-0387016P.	
PR	07-JUN-2002;	2002US-038796P.	
PR	07-JUN-2002;	2002US-038816P.	
PR	07-JUN-2002;	2002US-0388931P.	
PR	07-JUN-2002;	2002US-038942P.	
PR	07-JUN-2002;	2002US-038971P.	
PR	07-JUN-2002;	2002US-0387262P.	
PR	08-JUN-2002;	2002US-0296960P.	
PR	10-JUN-2002;	2002US-0387400P.	
PR	10-JUN-2002;	2002US-0387535P.	
PR	11-JUN-2002;	2002US-0387610P.	
PR	11-JUN-2002;	2002US-0387625P.	
PR	11-JUN-2002;	2002US-0387634P.	
PR	11-JUN-2002;	2002US-0387668P.	
PR	11-JUN-2002;	2002US-0387686P.	
PR	11-JUN-2002;	2002US-0387702P.	
PR	11-JUN-2002;	2002US-0387816P.	
PR	11-JUN-2002;	2002US-0387933P.	
PR	12-JUN-2002;	2002US-0387934P.	
PR	12-JUN-2002;	2002US-0387960P.	
PR	12-JUN-2002;	2002US-0388022P.	
PR	12-JUN-2002;	2002US-0388096P.	
PR	13-JUN-2002;	2002US-0389123P.	
PR	14-JUN-2002;	2002US-0389118P.	
PR	14-JUN-2002;	2002US-0389120P.	
PR	14-JUN-2002;	2002US-0389144P.	
PR	14-JUN-2002;	2002US-0389146P.	
PR	17-JUN-2002;	2002US-0389729P.	
PR	17-JUN-2002;	2002US-0389742P.	
PR	18-JUN-2002;	2002US-0389884P.	
PR	19-JUN-2002;	2002US-0390006P.	
PR	19-JUN-2002;	2002US-0390209P.	
PR	21-JUN-2002;	2002US-0390763P.	
XX			
PR	17-JUN-2002;	2002US-0396706P.	
PR	06-AUG-2002;	2002US-0401628P.	
PR	09-AUG-2002;	2002US-0402156P.	
PR	09-AUG-2002;	2002US-0402256P.	
PR	09-AUG-2002;	2002US-0402389P.	
PR	12-AUG-2002;	2002US-0402786P.	
PR	12-AUG-2002;	2002US-0402816P.	
PR	12-AUG-2002;	2002US-0402821P.	
PR	12-AUG-2002;	2002US-0402832P.	
PR	13-AUG-2002;	2002US-0403448P.	
PR	13-AUG-2002;	2002US-0403459P.	
PR	13-AUG-2002;	2002US-0403531P.	
PR	13-AUG-2002;	2002US-0403532P.	
PR	13-AUG-2002;	2002US-0403563P.	
PR	13-AUG-2002;	2002US-0406317P.	
PR	15-AUG-2002;	2002US-0406361P.	
PR	26-AUG-2002;	2002US-0406182P.	
PR	26-AUG-2002;	2002US-0406355P.	
PR	27-AUG-2002;	2002US-0406240P.	
PR	12-SEP-2002;	2002US-0410084P.	
PR	20-SEP-2002;	2002US-041258P.	
PR	23-SEP-2002;	2002US-0412711P.	
PR	30-SEP-2002;	2002US-0414801P.	
PR	30-SEP-2002;	2002US-0414839P.	
PR	30-SEP-2002;	2002US-0414840P.	
PR	30-SEP-2002;	2002US-0414954P.	</

SQ		Sequence 3546 AA:	
Query Match		94.5%; Score 16057; DB 8; Length 3546;	
Best Local Similarity		96.7%; Pred. No. 0;	
Matches 2956; Conservative 14; Mismatches 49; Indels 38; Gaps 10;			
Qy	2	AGAPPAL-----LPCSLISDCCASNORHSVGVSPSELVKKQIELKRGVLMSSKNSQ	57
Db	249	ASLAPVIVSSKMWLIHFTSD---GNHRQR-GFSAQIVQKKQIELKRGVLMSSKNSQ	304
Qy	58	KTSVLTGVGVSGQHNMCDPGI PERGKRLGSDPRLGSSVQFTCEGYDLOGSKRITCMY	117
Db	305	KTSVLTGVGVSGQHNMCDPGI PERGKRLGSDPRLGSSVQFTCEGYDLOGSKRITCMY	364
Qy	118	SDMFAASDHRPVCRAHMCDAHLRPGSIIISPMFPIQYDNNAICWIIITALNPSKIKL	177
Db	365	SDMFAASDHRPVCRAHMCDAHLRPGSIIISPMFPIQYDNNAICWIIITALNPSKIKL	424
Qy	178	AFEEFDLERGYDILTVDGGGDDQDKTVLYMSQVACSDSPHTPGSRIIPESMSGDIWRQKW	237
Db	425	AFEEFDLERGYDILTVDGGGDDQDKTVLYM-----LIGTSVVDLIVSTN-HQMW	473
Qy	238	TVLEICDIDISSDARSQSVKSPKTSNAVELVAPGTEIEQSGCGDPGIPAYGRREGSRFH	297
Db	474	LLFG-----TDG-SGS-----SLGPKASYEIEIEQSGCGDPGIPAYGRREGSRFR	516
Qy	298	HGDTLKEECOPAFELVQKAKITCOXNNQMSAKKRGCVFSCFENFTSPSGVULSPNYEDY	357
Db	517	HGDTLKEECOPAFELVQKAKITCOXNNQMSAKKRGCVFSCFENFTSPSGVULSPNYEDY	576
Qy	358	GNHLHCWLIILARBERSIHLAFNDIDIVEPQDFELVIKDGTABAPVLGTFSGNQDPSIIT	417
Db	577	GNHLHCWLIILARBERSIHLAFNDIDIVEPQDFELVIKDGTABAPVLGTFSGNQDPSIIT	636
Qy	418	SSGVARLEECOTHSSTGKRGPNITFTFRHNECPDGPVNGKRPDSDLOGSISIFLCD	477
Db	637	SSGVARLEECOTHSSTGKRGPNITFTFRHNECPDGPVNGKRPDSDLOGSISIFLCD	696
Qy	478	EGFLGTGSEBITTCVLEKGSVWNNSAVLRCEAPCGGHLTSPSGTILSPGMPGYKDALSC	537
Db	697	EGFLGTGSEBITTCVLEKGSVWNNSAVLRCEAPCGGHLTSPSGTILSPGMPGYKDALSC	756
Qy	538	AMVIEAOPGYPIKITDRFKTEVNYDTLEVDRGRITVSAPIIGVHGTQVPOFLISTSNYL	597
Db	757	AMVIEAOPGYPIKITDRFKTEVNYDTLEVDRGRITVSAPIIGVHGTQVPOFLISTSNYL	816
Qy	598	YLLFTSDKSHSDIGFOLRYETITLOSCHLDPGI PVNGQRHGDVFGALJNTPSCDSGT	657
Db	817	YLLFTSDKSHSDIGFOLRYETITLOSCHLDPGI PVNGQRHGDVFGALJNTPSCDSGT	876
Qy	658	LSDGEPLCECPNFQMSRALPSCCALCGFIOGSSGTILSPGFPDYPNNINACTWIIETSH	717
Db	877	LSDGEPLCECPNFQMSRALPSCCALCGFIOGSSGTILSPGFPDYPNNINACTWIIETSH	936
Qy	718	GKGVFTFHTFHESHGDVILITENGSTFOPLRQLTGSRLPAPISAGLYGNFTTAOVRFS	777
Db	937	GKGVFTFHTFHESHGDVILITENGSTFOPLRQLTGSRLPAPISAGLYGNFTTAOVRFS	996
Qy	778	DFSMAYEGFNITFSEYDLEPCEPEVPAYSIRKGLQFGVGDILTFCSPFGYLEGGART	837
Db	997	DFSMAYEGFNITFSEYDLEPCEPEVPAYSIRKGLQFGVGDILTFCSPFGYLEGGART	1056
Qy	838	CLGGRRLMSSPLPRVCAECGNSVTGTQGTLLSPNFVANNNNHCTIYSIQTPQKGIQL	897
Db	1057	CLGGRRLMSSPLPRVCAECGNSVTGTQGTLLSPNFVANNNNHCTIYSIQTPQKGIQL	1116
Qy	898	KARAFELSEGVDLYKYVDGNNNSARLIGVSHSEMMGVTLNASTSSLMDFITDANTSKG	957
Db	1117	KARAFELSEGVDLYKYVDGNNNSARLIGVSHSEMMGVTLNASTSSLMDFITDANTSKG	1176
Qy	958	FELHFSFELIKCEDPGTKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSBELLCISGRR	1017
Db	1177	FELHFSFELIKCEDPGTKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSBELLCISGRR	1236

Qy	1018	TMDRPLPTCVABCGGTVRGEVSGOVLSPGYPAPAYEHNNICITIEAAGCTTIGLHFLVPD	1077
Db	1237	TMDRPLPTCVABCGGTVRGEVSGOVLSPGYPAPAYEHNNICITIEAAGCTTIGLHFLVPD	1296
Qy	1078	TERVHDLVLRINQPVESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFFTSKQFALQFS	1137
Db	1297	TERVHDLVLRINQPVESGVLLKELSGPALPKDLHSTFNSVVLQFSTDFFTSKQFALQFS	1356
Qy	1138	VSTATSCNDPGIPONGSRGSDSMEAGDSVFOCDPGVALQGAELISCVKTEHFRFMOPBP	1197
Db	1357	GSTATSCNDPGIPONGSRGSDSMEAGDSVFOCDPGVALQGAELISCVKTEHFRFMOPBP	1416
Qy	1198	PTCIAPCGGDLTGSPGVILSPNYPBPYKCEKDMKVTSPDYIALV-FNIFNLEPGYD	1256
Db	1417	PTCIAPCGGDLTGSPGVILSPNYPBPYKCEKDMKVTSPDYIALV-FNIFNLEPGYD	1476
Qy	1257	FLHITDGRDLSPLIGSFYGSQLPGRITSSSNSLFLARSDASVSNAGFVIDFTNPRBS	1316
Db	1477	FLHITDGRDLSPLIGSFYGSQLPGRITSSSNSLFLARSDASVSNAGFVIDFTNPRBS	1536
Qy	1317	CFDPGSINKGTRVSGDLKLGSSVTYYCHGAYEVEGTSTLSCILGPDGKPRVNNPRPVCTA	1376
Db	1537	CFDPGSINKGTRVSGDLKLGSSVTYYCHGAYEVEGTSTLSCILGPDGKPRVNNPRPVCTA	1596
Qy	1377	PCGGQVSGSDGVYLSPNYPONTSGQICLYVTVVPKDY-VVFGQPAFHTALNDVEVHD	1435
Db	1597	PCGGQVSGSDGVYLSPNYPONTSGQICLYVTVVPKDY-VVFGQPAFHTALNDVEVHD	1656
Qy	1436	GHSCSRLLSSLSGHTGESLPLATSNQVLLKFSKGLAPARGHFVYQVAPRISATQCS	1495
Db	1657	GHSCSRLLSSLSGHTGESLPLATSNQVLLKFSKGLAPARGHFVYQVAPRISATQCS	1716
Qy	1496	SYPEPRYKRLGSDPSVGAIVRFECNSGVALQGSPEIECLVPGALQMNVSAPTCVPC	1555
Db	1717	SYPEPRYKRLGSDPSVGAIVRFECNSGVALQGSPEIECLVPGALQMNVSAPTCVPC	1776
Qy	1556	GGNLTERRGTLSPGFPPEPYNLSNCWKIVVPEGAGIQIOVVSFVTEQWMDSLLEVDA	1615
Db	1777	GGNLTERRGTLSPGFPPEPYNLSNCWKIVVPEGAGIQIOVVSFVTEQWMDSLLEVDA	1836
Qy	1616	DNTVYMLGSPGTVPPALNLSNSTNQLYLHFFSDISVSAAGFHLKYKTVGSSCEPAVPS	1675
Db	1837	DNTVYMLGSPGTVPPALNLSNSTNQLYLHFFSDISVSAAGFHLKYKTVGSSCEPAVPS	1896
Qy	1676	NGVKTGERYLVNDVVSFOCEBEGVALQGHANISCMPTGVRANNYPPLCIAQCGGTVBEM	1735
Db	1897	NGVKTGERYLVNDVVSFOCEBEGVALQGHANISCMPTGVRANNYPPLCIAQCGGTVBEM	1956
Qy	1736	GVLSPGPPGNYPPNNMCSMKIALPVGFGAHIOFLNSTEPNHDYIEIRNGPYETSRMG	1795
Db	1957	GVLSPGPPGNYPPNNMCSMKIALPVGFGAHIOFLNSTEPNHDYIEIRNGPYETSRMG	2016
Qy	1796	RFGSSELPSLSLSTSHETTYVPHSDSQNRPGFLXEQVAVLEQCPDPEPFANGIVAGAG	1855
Db	2017	RFGSSELPSLSLSTSHETTYVPHSDSQNRPGFLXEQVAVLEQCPDPEPFANGIVAGAG	2076
Qy	1856	YVNGOSVTFECLPQYOUTLGHVPULTCOHGTNRMDHPLPKCEVRCGNTITSNGTVSPGF	1915
Db	2077	YVNGOSVTFECLPQYOUTLGHVPULTCOHGTNRMDHPLPKCEVRCGNTITSNGTVSPGF	2136
Qy	1916	PSPYSSQDCVWLITVYIGHGVRNLNLSLQTESGDPFITIMDPPOQAPAPLGVFTSMK	1975
Db	2137	PSPYSSQDCVWLITVYIGHGVRNLNLSLQTESGDPFITIMDPPOQAPAPLGVFTSMK	2196
Qy	1976	KTYOSSNOVLKFRHDAATGIFAIAFSAYPLTKCPPTIILPNAEVTENEBFNIGDIY	2035
Db	2197	KTYOSSNOVLKFRHDAATGIFAIAFSAYPLTKCPPTIILPNAEVTENEBFNIGDIY	2256
Qy	2036	RYKCLPGFTLVGNELLTCKLGTVLOFEGPPICEVHCPTNELLTDSGTVILSSYSGSY	2095
Db	2257	RYKCLPGFTLVGNELLTCKLGTVLOFEGPPICEVHCPTNELLTDSGTVILSSYSGSY	2316

PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-0412731P.
 PR 30-SEP-2002; 2002US-0414801P.
 PR 30-SEP-2002; 2002US-0414839P.
 PR 30-SEP-2002; 2002US-0414840P.
 PR 30-SEP-2002; 2002US-0414954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-0423796P.
 PR 12-NOV-2002; 2002US-0425453P.
 (CURA-) CURAGEN CORP.
 XX
 PA
 PI Alsbrook JP, Alvarez E, Anderson DM, Boldog FI, Caeman SJ,
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,
 PI Etenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X,
 PI Guev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,
 PI MacLachlan T, Malyanekar UM, Mexick AJ, Millet I, Mishra VS,
 PI Pacteraju M, Paturajan M, Pena CE, Peyman SA, Raha D, Rastelli L,
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkels RA,
 PI Smithson G, Spytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M,
 PI Zhong H;
 XX
 DR WPI; 2004-081935/08.
 DR N-PSDB; ADH71143.
 XX
 PT New NOXV polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOXV-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 4; SEQ ID NO 40; 1880bp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (NOXV). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antihypertensive activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOXV polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOXV polypeptide of the invention.
 XX
 SO Sequence 3483 AA;
 Query Match 92.2%; Score 15655.5; DB 8; Length 3483;
 Best Local Similarity 94.6%; Pred. No. 0;
 Matches 2893; Conservative 14; Mismatches 49; Indels 101; Gaps 11;
 QY 2 AAGPPAL-----LPPSLISDCCASQNRHSVGVSELVKQIEKSRGVKLMPKDNQ 57
 DB 249 ASLPAAVSSIKWMLRLHFTSD--GNHROR-GFSAQYQVKQIEKSRGVKLMPKDNQ 304
 QY 58 KTSVLTQVGVSOQHNNCPDGGI PERGKRIGSDPFRGSSVQTFNCGYLOQSKRTTCMKV 117
 DB 305 KTSVLTQVGVSOQHNNCPDGGI PERGKRIGSDPFRGSSVQTFNCGYLOQSKRTTCMKV 364
 QY 118 SGMFAAMSDHPRVRCARMCDALRGPSGIITSPNPIQYDNNAHCWMTITLNPISKVTL 177

DB 365 SGMFAAMSDHPRVRCARMCDALRGPSGIITSPNPIQYDNNAHCWMTITLNPISKVTL 424
 QY 178 APEEPDLERGYDTLTVGDGDDGDDQKTVLYMGNQNCSDSPHTPGSRIPESMGDIWROKM 237
 DB 425 APEEPDLERGYDTLTVGDGDDGDDQKTVLYM-----LITGSRVPLIYSTN-HQNM 473
 QY 238 TVLEICRDISSDANSGRVRSKPKTSNAVELVAPGTEIEQSGCGDPGIPAYRRREGSRPH 297
 DB 474 LLEQ-----TDG-SGS-----SLGFASYSIEIEQSGCGDPGIPAYRRREGSRPH 516
 QY 298 HSDTLKFCOPAFELVQKAITCOKNOMSAKPCGVSCFENFTSPGVULSPVYPRY 357
 DB 517 HSDTLKFCOPAFELVQKAITCOKNOMSAKPCGVSCFENFTSPGVULSPVYPRY 576
 QY 358 GNLHCWMLILARPSRIHLAFNDIDVEPODFLYIKDGTAEABVLGTFSGNQPLSSIT 417
 DB 577 GNLHCWMLILARPSRIHLAFNDIDVEPODFLYIKDGTAEABVLGTFSGNQPLSSIT 636
 QY 418 SSGHVARLEFQDHSHTGKRGFNITTTFRANHCEDPVPVNGKRGDSLOQSSISFLCD 477
 DB 637 SSGHVARLEFQDHSHTGKRGFNITTTFRANHCEDPVPVNGKRGDSLOQSSISFLCD 696
 QY 478 EGFLEGTQSEITTCVAKESGVVWNSAVLRCEAPCGHLTSPSGTILSPGWGPFYDALSC 537
 DB 697 EGFLEGTQSEITTCVAKESGVVWNSAVLRCEAPCGHLTSPSGTILSPGWGPFYDALSC 756
 QY 538 AMVIAQGYPIKITEPDKTEVNDILEVRDRTYSAPLIGYHGTQVPOPLISTSNYL 597
 DB 757 AMVIAQGYPIKITEPDKTEVNDILEVRDRTYSAPLIGYHGTQVPOPLISTSNYL 816
 QY 598 YLLFSTDSSHSDIGQLREYETITLQSDHCLDPGI PVNGQRHGNPFYVALTTPGDSGYT 657
 DB 817 YLLFSTDSSHSDIGQLREYETITLQSDHCLDPGI PVNGQRHGNPFYVALTTPGDSGYT 876
 QY 817 LSDGEPLCEBNPQWRSALPSCALCGGFIQSSGTLISPPGDPFYNLNCTWIIETSH 717
 DB 877 LSDGEPLCEBNPQWRSALPSCALCGGFIQSSGTLISPPGDPFYNLNCTWIIETSH 936
 QY 718 GKGVEFTFHTHLSGHDYLLITENGSTFQPLRLQTSRLPAPISAGLYGNFTAOVRPIS 777
 DB 937 GKGVEFTFHTHLSGHDYLLITENGSTFQPLRLQTSRLPAPISAGLYGNFTAOVRPIS 996
 QY 778 DFSMSYEGFNITFSEYDLEPCSEBEVPAYSIIRKLOQGVGDTLFPSCFPGRLEGTRIT 837
 DB 997 DFSMSYEGFNITFSEYDLEPCSEBEVPAYSIIRKLOQGVGDTLFPSCFPGRLEGTRIT 1056
 QY 838 CLGRRRLMSPLPRCAECGNSVTGTGTLISNFPVNNNHCTIYSIQTOGKGIOL 897
 DB 1057 CLGRRRLMSPLPRCAECGNSVTGTGTLISNFPVNNNHCTIYSIQTOGKGIOL 1116
 QY 898 KARAFELSEGVLKVYDGNNSARLIGVPSHEMKGVTLNSTGSSMLDFTTDAENTSKG 957
 DB 1117 KARAFELSEGVLKVYDGNNSARLIGVPSHEMKGVTLNSTGSSMLDFTTDAENTSKG 1176
 QY 958 FELHFSSEFELIKCEDPPTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSEELCLSGERR 1017
 DB 1177 FELHFSSEFELIKCEDPPTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGSEELCLSGERR 1236
 QY 1018 TMDRPLPTCAECCGTVRGEVSGVLSPGYPAEYHNLCIWTIEAAGCTTGLHFLVPD 1077
 DB 1237 TMDRPLPTCAECCGTVRGEVSGVLSPGYPAEYHNLCIWTIEAAGCTTGLHFLVPD 1296
 QY 1078 TEEVHDVLRITMDPVEVSGVLLKELSGPALPKDLSHTFNSVVLQFSTDPFSSKGFALQPS 1137
 DB 1297 TEEVHDVLRITMDPVEVSGVLLKELSGPALPKDLSHTFNSVVLQFSTDPFSSKGFALQPS 1356
 QY 1138 VSTATSCNDPGLTQNSRSGDSWEAGDSVTFQCDPGYALQGSALISCVKIEENFFMQPSP 1197
 DB 1357 GSTATSCNDPGLTQNSRSGDSWEAGDSVTFQCDPGYALQGSALISCVKIEENFFMQPSP 1416
 QY 1198 PTCIAPCGDLTQPSGVILSPNPEPYPRGKEDCMKVTYSPDYVILV-FNIFNLEPGYD 1256
 DB 1417 PTCIAPCGDLTQPSGVILSPNPEPYPRGKEDCMKVTYSPDYVILV-FNIFNLEPGYD 1476

QY 1257 FLHIYDRDLSPLIGSFYGSOLPGRIESSNSLFLA PRSDASVSNAGFVTDYENRES 1316
 DB 1477 FLHIYDRDLSPLIGSFYGSOLPGRIESSNSLFLA PRSDASVSNAGFVTDYENRES 1536
 QY 1317 CFPDGS1KNGTRVSDKLGSSVYTYCHGVEEGTSLSCILGPDGKPVMMNRPVCTA 1376
 DB 1537 CFPDGS1KNGTRVSDKLGSSVYTYCHGVEEGTSLSCILGPDGKPVMMNRPVCTA 1596
 QY 1377 PCGGQYVSGDGVLSPVYPNQYISGQICLYFTVYPKDY - VFGQPAFPTALNDVEVD 1435
 DB 1597 PCGGQYVSGDGVLSPVYPNQYISGQICLYFTVYPKDYVFCQFAFPTALNDVEVD 1656
 QY 1436 GHSHSRLLSLSGSHGESLPLATSNQVLIKFSAKGLAPRGHFFHYQVAPRTSAQCS 1495
 DB 1657 GHSHSRLLSLSGSHGESLPLATSNQVLIKFSAKGLAPRGHFFHYQVAPRTSAQCS 1716
 QY 1496 SVPEPRYKRLGSDFSVCAIVREFCNSGYALQGSPEIECLPVGALAQMNVSAFTCVPC 1555
 DB 1717 SVPEPRYKRLGSDFSVCAIVREFCNSGYALQGSPEIECLPVGALAQMNVSAFTCVPC 1776
 QY 1556 GGNITERGTTLSGFPBPYINSNCWKIVPEAGIQIOVVSFVTEQWMDSLFVFGA 1615
 DB 1777 GGNITERGTTLSGFPBPYINSNCWKIVPEAGIQIOVVSFVTEQWMDSLFVFGA 1836
 QY 1616 DNTVTMLGSPGCTVPALLNSTSNQYLHFYSDSVSAAGFHEKTVGLSGCEPAVPS 1675
 DB 1837 DNTVTMLGSPGCTVPALLNSTSNQYLHFYSDSVSAAGFHEKTVGLSGCEPAVPS 1896
 QY 1676 NGVATGERVYLVNDVVSFCQCEGYALQGHAIHSCMPGTVRMMYPPCLIAQCGTVEME 1735
 DB 1897 NGVATGERVYLVNDVVSFCQCEGYALQGHAIHSCMPGTVRMMYPPCLIAQCGTVEME 1956
 QY 1736 GVILSPGPBGNYPSNMDCSWKIALPVGFAHIQFLNSTEBNHYEIRNGPYETSMMG 1795
 DB 1957 GVILSPGPBGNYPSNMDCSWKIALPVGFAHIQFLNSTEBNHYEIRNGPYETSMMG 2016
 QY 1796 RFGSSELPSSLISHTETVYFHSDHSONRGFLXEOAYELOCPPEPANIVGAG 1855
 DB 2017 RFGSSELPSSLISHTETVYFHSDHSONRGFLXEOAYELOCPPEPANIVGAG 2076
 QY 1856 YNVQGSVTFECLPGYQLTGHFVLTCQHTNRNMHPLPKCEVPCGNTITSNGTVSPGF 1915
 DB 2077 YNVQGSVTFECLPGYQLTGHFVLTCQHTNRNMHPLPKCEVPCGNTITSNGTVSPGF 2136
 QY 1916 PPSYSSQDCVWLITVPIGHGVRLNLSLQTEPSGDFITWDPQOTAPRLGVFTSMAX 1975
 DB 2137 PPSYSSQDCVWLITVPIGHGVRLNLSLQTEPSGDFITWDPQOTAPRLGVFTSMAX 2196
 QY 1976 KTVQSSNOVLKTKHRAATGCIPIAIAFSAVPLTKCPRPTLLPAAEVVTEBENIDIV 2035
 DB 2197 KTVQSSNOVLKTKHRAATGCIPIAIAFSAVPLTKCPRPTLLPAAEVVTEBENIDIV 2256
 QY 2036 RYRCLPFTLVGNELITCKLTGYLQFEGAPPICVNHCPTEMLTDSYGVILSOSYPSYP 2095
 DB 2257 RYRCLPFTLVGNELITCKLTGYLQFEGAPPICVNHCPTEMLTDSYGVILSOSYPSYP 2316
 QY 2096 QFQTCMWLVREPDYNI SLTYEYFLSEKQYDEFEIPDPSQSPBLKALSGNYSAPLIVT 2155
 DB 2317 QFQTCMWLVREPDYNI SLTYEYFLSEKQYDEFEIPDPSQSPBLKALSGNYSAPLIVT 2376
 QY 2156 SSSNSVYLRMSDDAANYRKGFKIYSAFYCSLPAPLHGFILQGTSTQPGSHFGCNAQ 2215
 DB 2377 SSSNSVYLRMSDDAANYRKGFKIYSAFYCSLPAPLHGFILQGTSTQPGSHFGCNAQ 2402
 QY 2216 YRLVGHSMALCTRRHPOGYHLSSEAIPLCOALSGLPAPKNGMFGKEVYTGTAVYSCS 2275
 DB 2403 -----ALSCGLPAPKNGMFGKEVYTGTAVYSCS 2433
 QY 2276 EGYHLQAGAEATACLDLTGLMSNNVPPQCVPTCPDVSSISVEHGMRLIFETOYQFOA 2335
 DB 2434 EGYHLQAGAEATACLDLTGLMSNNVPPQCVPTCPDVSSISVEHGMRLIFETOYQFOA 2493

QY 2336 QLMILCDPGYYTYGQVIRCOANGKMSLGDSTPTCRILISGCELPPIPNHGRICTLASYGA 2395
 DB 2494 QLMILCDPGYYTYGQVIRCOANGKMSLGDSTPTCRILISGCELPPIPNHGRICTLASYGA 2553
 QY 2396 TAI FSCNSGYTLVGSVRRECMAWLGSGSEVRCLAGHCGTPEBIVNGHGENYVXGSV 2455
 DB 2554 TAI FSCNSGYTLVGSVRRECMAWLGSGSEVRCLAGHCGTPEBIVNGHGENYVXGSV 2613
 QY 2456 VYQCNAGFLIGMSVRI COODHMSGKTPEVCVITTCGHPNPVNGLTQGNQFNLNDVYKF 2515
 DB 2614 VYQCNAGFLIGMSVRI COODHMSGKTPEVCVITTCGHPNPVNGLTQGNQFNLNDVYKF 2673
 QY 2516 VCNPGVABGAARSOCIASGQMSDMLPTCII INCTPGHOENSVRQVHAGSPRFSCTT 2575
 DB 2674 VCNPGVABGAARSOCIASGQMSDMLPTCII INCTPGHOENSVRQVHAGSPRFSCTT 2733
 QY 2576 VSYRCHNFFYLLGTPLYLSCQGDGTWDRPREOCILVSCGHPSPHSQMSGDSYTVGAVVR 2635
 DB 2734 VSYRCHNFFYLLGTPLYLSCQGDGTWDRPREOCILVSCGHPSPHSQMSGDSYTVGAVVR 2793
 QY 2636 YSCIGKRTLVGNSTRMCGLDGHWGSLPHCSGTSVYCGDPGIPANGIRLGDSPDPTVM 2695
 DB 2794 YSCIGKRTLVGNSTRMCGLDGHWGSLPHCSGTSVYCGDPGIPANGIRLGDSPDPTVM 2853
 QY 2696 RFGCEAGHYLRGSSERTCOANGSMGSGQPECGVISCNPPTPSNARVVFEDGLVFSSIV 2755
 DB 2854 RFGCEAGHYLRGSSERTCOANGSMGSGQPECGVISCNPPTPSNARVVFEDGLVFSSIV 2913
 QY 2756 YECREGYATGLSRHCSVNGVTGSDPECLVINCQDPGIPANGIRLGDSPDPTVM 2815
 DB 2914 YECREGYATGLSRHCSVNGVTGSDPECLVINCQDPGIPANGIRLGDSPDPTVM 2973
 QY 2816 CVPGYMESHRSVYLSCTKRTYVNGTKPVCKALMCKRPPILPBGKVVGSDFPMGSSVTYA 2875
 DB 2974 CVPGYMESHRSVYLSCTKRTYVNGTKPVCKALMCKRPPILPBGKVVGSDFPMGSSVTYA 3033
 QY 2876 CLBGYQLSLPAVFTCEGNSWTEGELPQCFVFGCDPVPRGRREDGFSYRSVSFSCH 2935
 DB 3034 CLBGYQLSLPAVFTCEGNSWTEGELPQCFVFGCDPVPRGRREDGFSYRSVSFSCH 3093
 QY 2936 PPLVLVGSPPRFQSDGTWSTGTPSCIDPTLTTCADPGVBOFGIUNNSQGYOVGSTVLF 2995
 DB 3094 PPLVLVGSPPRFQSDGTWSTGTPSCIDPTLTTCADPGVBOFGIUNNSQGYOVGSTVLF 3153
 QY 2996 COKGYLLQGSTTTCPLPNTLWSTGTPDPCVHRCROBETPTHANVAGALDPSMGYTLI 3052
 DB 3154 COKGYLLQGSTTTCPLPNTLWSTGTPDPCVHRCROBETPTHANVAGALDPSMGYTLI 3210

RESULT 7
 ADH71146
 ID ADH71146 standard; protein; 3130 AA.
 AC ADH71146;
 XX
 DT 25-MAR-2004 (first entry)
 XX
 DE Human protein of the invention NOVAE SEQ ID NO:42.
 XX
 KW human; cytosolic; immunomodulator; neuroprotective; nootropic;
 KW anorectic; antidiabetic; antimicrobial; antiparasitic; gene therapy;
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
 KW obesity; diabetes; infectious disease; metabolic syndrome X;
 KW dyslipidemia.
 XX
 OS Homo sapiens.
 XX
 PN NO2003102155-A2.
 XX
 PD 11-DEC-2003.
 XX
 PF 03-JUN-2003; 2003WO-US017430.
 XX

PR	05-NOV-2002; 2002JUS-00423798.	
PR	05-NOV-2002; 2002JUS-00423798P.	
PR	12-NOV-2002; 2002JUS-0425453P.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Alcobrook JP, Alvarez E, Anderson DW, Boldog FI, Casman SJ,	
PI	Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,	
PI	Ettenberg S, Gangoli EA, Getchell VL, Gotman L, Gunther E, Guo X;	
PI	Gusev VI, Herrman JL, Ji W, Kakuda R, Li L, Liu X, MacDougall JR;	
PI	MacLachlan T, Malyanankar UM, Mezzick AJ, Millet I, Mishra VS;	
PI	Padiagaran M, Patturajan M, Pena CE, Peyman JA, Raha D, Rastelli L;	
PI	Rieger DK, Rothenberg ME, Score P, Shenoy SG, Shinkets RA;	
PI	Smithson G, Szytek KA, Stone DJ, Vernet CAM, Voss EZ, Zhong M;	
XX	Zhong H;	
DR		
DR	WPI; 2004-081935/08.	
XX	N-PSDB; ADH71145.	
XX		
PT	New NOXV polypeptides and nucleic acid molecules useful for preventing or	
PT	treating NOXV-associated disorders, e.g. cancer, diabetes, infection or	
PT	obesity, and in chromosome mapping, tissue typing or pharmacogenomics.	
XX		
XX	Example 4; SEQ ID NO 42; 1880bp; English.	
CC		
CC	The invention relates to a novel isolated polypeptide (NOXV). A	
CC	polypeptide of the invention has cytostatic, immunomodulator,	
CC	neuroprotective, nocotropic, anorectic, antidiabetic, antimicrobial, and	
CC	antihypaemic activity, and may have a use in gene therapy, and as a	
CC	vaccine. The polypeptides are encoded by NOXV polynucleotides comprising	
CC	any of the 303 fully defined nucleotide sequences given in the	
CC	specification. The polypeptide is useful in the manufacture of a	
CC	medicament for treating a syndrome associated with a human disease. The	
CC	polypeptide, polynucleotide and antibody are useful in diagnosing,	
CC	treating or preventing NOXV-associated disorders, e.g. cancer, cachexia,	
CC	Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious	
CC	diseases, metabolic syndrome X or dyslipidemias. The nucleic acids are	
CC	further used as hybridisation probes, in chromosome mapping, tissue	
CC	typing, preventive medicine, and pharmacogenomics. The present sequence	
CC	represents a NOXV polypeptide of the invention.	
XX		
SO	Sequence 3130 AA;	
	Query Match 89.6%; Score 15219; DB 8; Length 3130;	
	Best Local Similarity 93.5%; Pred. No. 0;	
	Matches 2812; Conservative 13; Mismatches 29; Indels 152; Gaps 9	
QY	50 MSKXNSQRTSVLTQVYGSQHNNCCPPGIPERKRGISGPRFGSSVQFTCNEGYLGS 109	
DB	1 MSKXNSQRTSVLTQVYGSQHNNCCPPGIPERKRGISGPRFGSSVQFTCNEGYLGS 60	
QY	110 KRITCMKTSDMFPAAMSDIRPVCARMCDALRGSGITLSPNPIQYDNNAHCVLITL 169	
DB	61 KRITCMKTSDMFPAAMSDIRPVCARMCDALRGSGITLSPNPIQYDNNAHCVLITL 120	
QY	170 NFSKVIKLAFFEPDLERGYDTLTVGGDGGDQDKTVLYMSQNAKSDSPHTPGSRIPESMS 229	
DB	121 NFSKVIKLAFFEPDLERGYDTLTVGGDGGDQDKTVLYM-----LTGNSVDDLIV 170	
QY	229 GDIMQKQTVLEICRDISSSABSGSVKPSKTSNAVELVAPETLEIGSSCGDPCGIPAVG 289	
DB	171 STM-HQMWLLPQ-----TDG-SGS-----SLGFKASVYEIEIGSSCGDPCGIPAVG 212	
QY	290 RREGSRFHGHGTLTFECOPAFELVGKAITCQKNNOMSAKKPGCVSPCFENFTSPSGVYL 349	
DB	213 RREGSRFHGHGTLTFECOPAFELVGKAITCQKNNOMSAKKPGCVSPCFENFTSPSGVYL 272	
QY	350 SPNIPEDYGNLHACVWLILARPESRIHLAENDIDVBPQDFLVIKDGATAEAPVLGTFSG 409	
DB	273 SPNIPEDYGNLHACVWLILARPESRIHLAENDIDVBPQDFLVIKDGATAEAPVLGTFSG 332	
QY	410 NOLBSITSSGHVARELFQNDHGTGKRGNTPTTTRHNRGCDPGVPPVNGKRGFGSLQ 469	

Db 333 NQLPSSITSSGHWARLEFQDHTSGKGFNITTSSESSNECPDPGVPVNGKRFQDLSLQ 392
 Qy 470 SSISFLDDEGLTQSGSETITCVLKEGSVVWNSAVLREAPCGHLSPSGCTILSPGMPG 529
 Db 393 SSISFLDDEGLTQSGSETITCVLKEGSVVWNSAVLREAPCGHLSPSGCTILSPGMPG 452
 Qy 530 FYKALSCAWIEAOPGPIKITFDRPFTEVNYDTLEVRDGRYTSAPLIGYHGTQVPOF 589
 Db 453 FYKALSCAWIEAOPGPIKITFDRPFTEVNYDTLEVRDGRYTSAPLIGYHGTQVPOF 512
 Qy 590 LISTSNLYLLFSTDKSHDIDGFLRYETITLQSDHCLDPGI PVNGQRHNDPVYALVT 649
 Db 513 LISTSNLYLLFSTDKSHDIDGFLRYETITLQSDHCLDPGI PVNGQRHNDPVYALVT 572
 Qy 650 FSCSGYTLSDGEPLCEPNFQMSBALPSCALCGGFIQSSSGTILSPGPFDPYNNLNC 709
 Db 573 FSCSGYTLSDGEPLCEPNFQMSBALPSCALCGGFIQSSSGTILSPGPFDPYNNLNC 632
 Qy 710 TWIETSHGKGVFTFHTFHLSEGHVLLITENGSPFOPLRQLTGSRLPAPISAGLYNF 769
 Db 633 TWIETSHGKGVFTFHTFHLSEGHVLLITENGSPFOPLRQLTGSRLPAPISAGLYNF 692
 Qy 770 TAOVRFISDFMSYEGFNITFSEYDLBECBEPVPAVSIRKGLQFGVGDILTFSQCPGXR 829
 Db 693 TAOVRFISDFMSYEGFNITFSEYDLBECBEPVPAVSIRKGLQFGVGDILTFSQCPGXR 752
 Qy 830 LEBGARITCLGRRRLWSSPLPRCYACGNSVTGTQTLSPNPVYNNNNHCTIYSIQ 889
 Db 753 LEBGARITCLGRRRLWSSPLPRCYACGNSVTGTQTLSPNPVYNNNNHCTIYSIQ 812
 Qy 890 QPGGIOLKARAFELSEGDVLYKVDGNNNSARLLGVSHSEMMCVTLNSTSSILMLPIT 949
 Db 813 QPGGIOLKARAFELSEGDVLYKVDGNNNSARLLGVSHSEMMCVTLNSTSSILMLPIT 872
 Qy 950 DAENTSKGFEHAFSSFELIKCEDPQTPKFGYKVHDEGHFAGSSVSPSCDPGYSLRGSEL 1009
 Db 873 DAENTSKGFEHAFSSFELIKCEDPQTPKFGYKVHDEGHFAGSSVSPSCDPGYSLRGSEL 932
 Qy 1010 LCLSGERTMDRPLPTCYAECGGTVRGVSGQVLSPGYPAHYENHNLCTITIEBAGCTI 1069
 Db 933 LCLSGERTMDRPLPTCYAECGGTVRGVSGQVLSPGYPAHYENHNLCTITIEBAGCTI 992
 Qy 1070 GLHFLVPDTEBVHVLRIWDPVBSGVLLKELSGPALPKDLHSTFNSVVLQFSTDPFTSK 1129
 Db 993 GLHFLVPDTEBVHVLRIWDPVBSGVLLKELSGPALPKDLHSTFNSVVLQFSTDPFTSK 1052
 Qy 1130 QGFALIOFSVSTATSCNDPGLPONGSRGSDSWEAGDSTVFQCDPGYALOGSAEISCXKEN 1189
 Db 1053 QGFALIOFSVSTATSCNDPGLPONGSRGSDSWEAGDSTVFQCDPGYALOGSAEISCXKEN 1112
 Qy 1190 RFPMQPSBPPTCIAQCGDLTPSPGVILSPNYPEPYPEGKECDMKVTVSPDVIILVFP 1248
 Db 1113 RFPMQPSBPPTCIAQCGDLTPSPGVILSPNYPEPYPEGKECDMKVTVSPDVIILVFP 1172
 Qy 1249 FNLPGVDLHIYVGRDLSPLIGSPYGSOLRGRIESSNSLFLAFPSDASVSNAGFTVD 1308
 Db 1173 FNLPGVDLHIYVGRDLSPLIGSPYGSOLRGRIESSNSLFLAFPSDASVSNAGFTVD 1232
 Qy 1309 YTENPRESCEPDGSIKNGTRVGSDLKLSSVTVYCHGGEVEGTSTLSCLIGPDGKPVWN 1368
 Db 1233 YTENPRESCEPDGSIKNGTRVGSDLKLSSVTVYCHGGEVEGTSTLSCLIGPDGKPVWN 1292
 Qy 1369 NPREVCTAPCGGQYVSDGVVLSRNYPONTSSGOI CLYFTVTPYQY-VVFGQAFEPHTAL 1427
 Db 1293 NPREVCTAPCGGQYVSDGVVLSRNYPONTSSGOI CLYFTVTPYQYVFGQAFEPHTAL 1352
 Qy 1428 NDVVEVMDHSHOHRLLSSLSGSHT-GBSLPLATSNQVLIKFSKAGLAPARGPFPYQAV 1486
 Db 1353 NDVVEVMDHSHOHRLLSSLSGSHTGBSLPLATSNQVLIKFSKAGLAPARGPFPYQAV 1412
 Qy 1487 PRISATOCSSVPEBRYGRLGSDFSVGAI VRFECNSGVALOGSPEIECLPVGALAOVNV 1546
 Db 1413 PRISATOCSSVPEBRYGRLGSDFSVGAI VRFECNSGVALOGSPEIECLPVGALAOVNV 1472

Qy 1547 SAPTCVPPCGGNLTERRGITLSPGPEBPYVNSLNCWKIIVPBGAGIOIOVSPFTEBOM 1606
 Db 1473 SAPTCVPPCGGNLTERRGITLSPGPEBPYVNSLNCWKIIVPBGAGIOIOVSPFTEBOM 1532
 Qy 1607 DSLEVPDADNTVYMLGSPSGTIVPALNINSTQVLYLHFPSDISVSNAGFHELEKTVGLS 1666
 Db 1533 DSLEVPDADNTVYMLGSPSGTIVPALNINSTQVLYLHFPSDISVSNAGFHELEKTVGLS 1592
 Qy 1667 SCPEPAVNSGVKTGERVYLVNDVVSFOCEBQYALQGHAIISCMGTVRRNNYPPPLCIAQ 1726
 Db 1593 SCPEPAVNSGVKTGERVYLVNDVVSFOCEBQYALQGHAIISCMGTVRRNNYPPPLCIAQ 1652
 Qy 1727 CGGTVEEMEGVILSPGPRGYPVNSMDSKMIALPVGGAHIQPLNFTSEBNHDIETIRNG 1786
 Db 1653 CGGTVEEMEGVILSPGPRGYPVNSMDSKMIALPVGGAHIQPLNFTSEBNHDIETIRNG 1712
 Qy 1787 PYETSRMAGRFSGSELSSILSTSHETTVYFHSDSQNRPGFKLEYQAYELOECBPDEPF 1846
 Db 1713 PYETSRMAGRFSGSELSSILSTSHETTVYFHSDSQNRPGFKLEYQAYELOECBPDEPF 1772
 Qy 1847 ANGIVRGAGYNVQSVTFECLPGYQLTGHPVLTQHGNTNMDHPLPKCEVPCCGNTSS 1906
 Db 1773 ANGIVRGAGYNVQSVTFECLPGYQLTGHPVLTQHGNTNMDHPLPKCEVPCCGNTSS 1832
 Qy 1907 NGTVYSGRPSPPYSSODCVMILTPVIGHVRLNLSLLOTEBPSGDFITINDGPOQTA PRL 1966
 Db 1833 NGTVYSGRPSPPYSSODCVMILTPVIGHVRLNLSLLOTEBPSGDFITINDGPOQTA PRL 1892
 Qy 1967 GVFTSRMAKKTVOSSNQVLLKFRDAATGAI FALAFSAPLTKCPPTILPNAEVTEN 2026
 Db 1893 GVFTSRMAKKTVOSSNQVLLKFRDAATGAI FALAFSAPLTKCPPTILPNAEVTEN 1952
 Qy 2027 EEFNIGDIVRYRCLPGFTLVGNELTCKLGTVYLFQFEGPPICEVHCETNELTIDSTGVIL 2086
 Db 1953 EEFNIGDIVRYRCLPGFTLVGNELTCKLGTVYLFQFEGPPICEVHCETNELTIDSTGVIL 2012
 Qy 2087 SOSYRGSYPOFOCTSMVRYEPPDYNISLTYEFLSEKQVDEFEIPDPSQOSPLKXLSG 2146
 Db 2013 SOSYRGSYPOFOCTSMVRYEPPDYNISLTYEFLSEKQVDEFEIPDPSQOSPLKXLSG 2072
 Qy 2147 NYSAPLIVTSSNSVYLRMSDHAYNRKGRKIRASACSLPAPLHGLTIGOTSTPGG 2206
 Db 2073 NYSAPLIVTSSNSVYLRMSDHAYNRKGRKIRASACSLPAPLHGLTIGOTSTPGG 2132
 Qy 2207 SIHFGNAGYRLVGHSMALCTRRPQGYHLMSEALPLCOALSCGLPEAPKXGMVFGKEYTV 2266
 Db 2133 SIHFGNAGYRLVGHSMALCTRRPQGYHLMSEALPLCOALSCGLPEAPKXGMVFGKEYTV 2192
 Qy 2267 GTKAVYSCSGHYHQAQAEATBCLDPTGLMSNNRVPPQCVPVCPDVSSISVEHGRRLI 2326
 Db 2193 GTKAVYSCSGHYHQAQAEATBCLDPTGLMSNNRVPPQCVPVCPDVSSISVEHGRRLI 2252
 Qy 2327 FETQYQFOQMLICDPGYTYTGORVYRCQANGKMSLGBDTPTRITISCELEPIPPGHR 2386
 Db 2253 FETQYQFOQMLICDPGYTYTGORVYRCQANGKMSLGBDTPTRITISCELEPIPPGHR 2312
 Qy 2387 IGTLSYVGAATAFSCNSGYTLVSGRYRECMANGLMSGSEVRLCAGHGTEPIVNGHNG 2446
 Db 2313 IGTLSYVGAATAFSCNSGYTLVSGRYRECMANGLMSGSEVRLCAGHGTEPIVNGHNG 2372
 Qy 2447 ENYSYRGSVYVYQCNAGRLLIGMSVRIQODHMSGKTPFCVPIITCGHGNPVNGLTQGNQ 2506
 Db 2373 ENYSYRGSVYVYQCNAGRLLIGMSVRIQODHMSGKTPFCVPIITCGHGNPVNGLTQGNQ 2413
 Qy 2507 FNLNDVVKFVCPNPGYMAEGAARSQLASQMSDMLPTCRILINCTDPGHQENSVQVHASG 2566
 Db 2414 FNLNDVVKFVCPNPGYMAEGAARSQLASQMSDMLPTCRILINCTDPGHQENSVQVHASG 2413
 Qy 2567 PHRPSFGTTVSYRNCNMGFVLLGTPVLSQGDGTWDRBRPOCLLVSCGHPSGPHSONSGD 2626
 Db 2414 PHRPSFGTTVSYRNCNMGFVLLGTPVLSQGDGTWDRBRPOCLLVSCGHPSGPHSONSGD 2431

QY 2627 SYTVGAIVVYSCIGKRTLVGNSTRMCGLDGHWLGLSPHCSSGTSGVGVCGDPGPIPAHGTIRLG 2686
 DB 2432 SYTVGAIVVYSCIGKRTLVGNSTRMCGLDGHWLGLSPHCSSGTSGVGVCGDPGPIPAHGTIRLG 2491
 QY 2687 DSFPDPTVNRFCSEAGHVLRRGSSERTQANSGWSGSGPECGVISCANPQTPSNARVPSD 2746
 DB 2492 DSFPDPTVNRFCSEAGHVLRRGSSERTQANSGWSGSGPECGVISCANPQTPSNARVPSD 2551
 QY 2747 GLVFSSTIYECREGVATGTLSSRHCSVNGWTGSDPECLVINCSDPGIIPANGRLGNDF 2806
 DB 2552 GLVFSSTIYECREGVATGTLSSRHCSVNGWTGSDPECLVINCSDPGIIPANGRLGNDF 2611
 QY 2807 RINKVTYVQCVGVYMMESHRSVLSCTDRITNGTKPVCKALMCKPPLIPNGKVGSDF 2866
 DB 2612 RINKVTYVQCVGVYMMESHRSVLSCTDRITNGTKPVCKALMCKPPLIPNGKVGSDF 2671
 QY 2867 MWGSSVTVACLEGYOLSLPAVPTCGNGSNMGELPQCPVPGCGDPGVRSRRRRDRGFSY 2926
 DB 2672 MWGSSVTVACLEGYOLSLPAVPTCGNGSNMGELPQCPVPGCGDPGVRSRRRRDRGFSY 2731
 QY 2927 RSSVSFSCHPPLVLVGSRRFCQSDGTWGTGTOPSCIDPTLTTCADPGVPORGIONNGY 2986
 DB 2732 RSSVSFSCHPPLVLVGSRRFCQSDGTWGTGTOPSCIDPTLTTCADPGVPORGIONNGY 2791
 QY 2987 QVGSVTVLRCCQGYLLQGSTTRTCLPNTWGTGTPPDVCVPHCRQPTPTAHNVGALDLP 3046
 DB 2792 QVGSVTVLRCCQGYLLQGSTTRTCLPNTWGTGTPPDVCVPHCRQPTPTAHNVGALDLP 2851
 QY 3047 MGYTLI 3052
 DB 2852 MGYTLI 2857

RESULT 8
 ID ABG79169 standard; protein; 2669 AA.
 XX ABG79169;
 AC
 AC 15-NOV-2002 (first entry)
 DT
 XX
 DE Human cub and sushi domain containing protein #2.
 DE
 XX
 KW Human; NOX; neurodegenerative disease; Alzheimer's disease; anxiety;
 KW Parkinson's disease; Huntington's disease; neurological disorder;
 KW schizophrenia; manic depression; mental retardation; angina pectoris;
 KW cardiovascular disease; acute heart failure; myocardial infarction;
 KW muscular disease; muscular disorder; retinal disease; photoreception;
 KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;
 KW immunological disorder; inflammatory disease; immune disease; diabetes;
 KW bacterial infection; fungal infection; protozoal infection; obesity;
 KW viral infection; reproductive system disorder; metabolic disturbance;
 KW anorexia; wasting disorder; chronic disease; infectious disease;
 KW dyslipidaemia; cub; sushi; myelin; von Willebrand factor; kiellin;
 KW semaphorin; serine/threonine protein kinase; TGF-beta binding;
 KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;
 KW colloid-like 2; cysteine sulfenic acid decarboxylase.
 XX
 OS Homo sapiens.
 OS
 XX
 PN WO200264791-A2.
 XX
 PD 22-AUG-2002.
 PD
 PF 10-DEC-2001; 2001MO-US048369.
 PF
 XX 08-DEC-2000; 2000US-0254329P.
 PR 14-DEC-2000; 2000US-0255648P.
 PR 15-MAY-2001; 2001US-0291037P.
 PR 08-JUN-2001; 2001US-0297173P.
 PR 08-JUN-2001; 2001US-0309258P.
 PR 29-AUG-2001; 2001US-0315639P.
 PR 01-OCT-2001; 2001US-0326393P.

XX
 FA (CURA-) CURAGEN CORP.
 XX
 PI Alebrook JP, Anderson DM, Burgess CE, Boidog FL, Casman SJ,
 PI Coleman SP, Edinger SR, Ellerman K, Gerlach V, Gorman L, Grosse WM,
 PI Guo X, Herrmann III, Kekuda R, Lepley DM, Li L, Macdougall JR,
 PI Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkets RA,
 PI Sutherland G, Szytek KA, Stone DJ, Tchernev VT, Vernet CAM, Voss EZ,
 PI Zernhusen BD, Zhong H, Zhong M,
 XX
 DR WPI; 2002-643486/69.
 DR
 N-PSDB; ABS64376.
 XX
 PT New NOX polypeptides and polynucleotides useful for treating or
 PT preventing e.g. neurodegenerative diseases, neurological disorders,
 PT cardiovascular diseases, muscular diseases and disorders, or
 PT immunological diseases.
 XX
 PS
 XX
 Claim 1; Page 16-17; 299pp; English.

The present invention relates to new NOX polypeptides. The polypeptides, CC
 polynucleotides and antibodies are useful in the manufacture of a CC
 medicament for treating or preventing neurodegenerative diseases (e.g. CC
 Alzheimer's disease, Parkinson's disease, or Huntington's disease), CC
 neurological disorders (e.g. anxiety, schizophrenia, manic depression or CC
 mental retardation), cardiovascular disease (e.g. acute heart failure, or CC
 angina pectoris or myocardial infarction), muscular diseases and CC
 disorders, retinal diseases (including those involving photoreception, CC
 deafness and keratinisation disorders), cancer (e.g. ovarian cancer or CC
 melanoma), immunological disorders, inflammatory and immune diseases, CC
 bacterial, fungal, protozoal and viral infections, and reproductive CC
 system disorders. The proteins of the invention may be used to screen CC
 drugs or compounds that modulate the NOX protein activity or expression, CC
 as well as to treat disorders characterised by insufficient or excessive CC
 production of NOX protein or protein forms that have decreased or CC
 aberrant activity compared to NOX wild type protein, such as diabetes, CC
 obesity, metabolic disturbances associated with obesity, anorexia and CC
 wasting disorders associated with chronic diseases and various cancers, CC
 infectious diseases and various dyslipidaemias. The nucleic acid CC
 sequences of the invention may be used in chromosome mapping, identifying CC
 an individual from minute biological samples (tissue typing), and in CC
 forensic identification of a biological sample. The present amino acid CC
 sequence represents a NOX protein of the invention
 XX
 SQ Sequence 2669 AA;

Query Match 83.3%; Score 14142.5; DB 5; Length 2669;
 Best Local Similarity 97.7%; Pred. No. 0;
 Matches 2605; Conservative 1; Mismatches 2; Indels 59; Gaps 3;

QY 1 MAGAPPALLLPCLISDCCASNORSHVGVPSFLVKKOIELKSRGVKMPKSDNOKTS 60
 DB 1 MAGAPPALLLPCLISDCCASNORSHVGVPSFLVKKOIELKSRGVKMPKSDNOKTS 60
 QY 61 VLTQGVSGQHNNCPDPIPERGRKLGSDPRLGSSVQFTNBEVDLQGSKRITCMKVSDM 120
 DB 61 VLTQGVSGQHNNCPDPIPERGRKLGSDPRLGSSVQFTNBEVDLQGSKRITCMKVSDM 120
 QY 121 FAAMSDHRPVCRAHRCMDAHLRGSGITTSNPFPIQYDNNAHCWITITLALNPSKVITLAF 180
 DB 121 FAAMSDHRPVCRAHRCMDAHLRGSGITTSNPFPIQYDNNAHCWITITLALNPSKVITLAF 180
 QY 181 EFDLERYGDTLTVGGDGDQDKTVLYMSQNASDSDPHPTGSRIPESMGDIIROKMTVL 240
 DB 181 EFDLERYGDTLTVGGDGDQDKTVLYMSQNASDSDPHPTGSRIPESMGDIIROKMTVL 240
 QY 241 EICRDISSDARSAGSVKSPKTSNAYELVAPGTETIQSGCGDGPAPAYRRREGSRPHHD 300
 DB 241 EICRDISSDARSAGSVKSPKTSNAYELVAPGTETIQSGCGDGPAPAYRRREGSRPHHD 300
 QY 301 TLKFECPAPFELVGOVAITCOKNNOWSAKKPGVSCFNFTSPSGVULSPMYPDYGVH 360
 DB 301 TLKFECPAPFELVGOVAITCOKNNOWSAKKPGVSCFNFTSPSGVULSPMYPDYGVH 360

QY 361 LHCWMLIARBSRIHLAFNDIDVEPODFLVIKOGATAEAPVLGTFESGNOLPSSITSSG 420
 Db 361 LHCWMLIARBSRIHLAFNDIDVEPODFLVIKOGATAEAPVLGTFESGNOLPSSITSSG 420
 QY 421 HVARLEFQDHSSTGRGFINITFTTFRNECEPCDPGVNPKRFGDSLQJGSSISPLCDEGF 480
 Db 421 HVARLEFQDHSSTGRGFINITFTTFRNECEPCDPGVNPKRFGDSLQJGSSISPLCDEGF 480
 QY 481 LGTQGSSTITCVLKEGVSVMNSAVLRBAPCGGHLTSPSGTILSPGMPGYKALSCAMV 540
 Db 481 LGTQGSSTITCVLKEGVSVMNSAVLRBAPCGGHLTSPSGTILSPGMPGYKALSCAMV 540
 QY 541 IEAOPGPIKTPRFRFTEVVDLLEVRDGTYSAPLIGVYHGOVOPILISTSNLYL 600
 Db 541 IEAOPGPIKTPRFRFTEVVDLLEVRDGTYSAPLIGVYHGOVOPILISTSNLYL 600
 QY 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGIPVNGORHNDFYVGAIVTESCDSGYTLSD 660
 Db 601 FSTDKSHSDIGFOLRYETITLQSDHCLDPGIPVNGORHNDFYVGAIVTESCDSGYTLSD 660
 QY 661 GEPLECEPNFQMSRALSCEALCCGFIQSSGTTLSFGPDPFYNNINCTWIIETSHGK 720
 Db 661 GEPLECEPNFQMSRALSCEALCCGFIQSSGTTLSFGPDPFYNNINCTWIIETSHGK 720
 QY 721 VFFPHTFHLESCHDYLLITENSGFTQPLRQLTGSRLPAPISAGLYGNFTAOVRFTSDS 780
 Db 721 VFFPHTFHLESCHDYLLITENSGFTQPLRQLTGSRLPAPISAGLYGNFTAOVRFTSDS 780
 QY 781 MSYGFNITSEYDLBCEBEVPAVSIKGLQGVDTLTFSCFPGLRGTARITCLG 840
 Db 781 MSYGFNITSEYDLBCEBEVPAVSIKGLQGVDTLTFSCFPGLRGTARITCLG 840
 QY 841 GRRRLMSSPLRCAECGNSVTGTQTLSPNPNVNNNHECTIYSIQTOPGKIQOKAR 900
 Db 841 GRRRLMSSPLRCAECGNSVTGTQTLSPNPNVNNNHECTIYSIQTOPGKIQOKAR 900
 QY 901 AFELSEBDVLKVYGNNNNSARLLGVFHSHEMVGTLNSTSSLMADITTAENTSKFEL 960
 Db 901 AFELSEBDVLKVYGNNNNSARLLGVFHSHEMVGTLNSTSSLMADITTAENTSKFEL 960
 QY 961 HFSFELIKCEDPTPKFGYKVNDEGHPAGSSVSFSCDPGSLRGSBELCLSGERRTWD 1020
 Db 961 HFSFELIKCEDPTPKFGYKVNDEGHPAGSSVSFSCDPGSLRGSBELCLSGERRTWD 1020
 QY 1021 RPLPTCAECGGTVRGEVSGVLSPGYPARYENHNLCTIWTIEBAGCTTGLHFLVDTSE 1080
 Db 1021 RPLPTCAECGGTVRGEVSGVLSPGYPARYENHNLCTIWTIEBAGCTTGLHFLVDTSE 1080
 QY 1081 VHDVLRITWDGVEBEGVLLKELSGPALPKDLHSTFNSVVLQSTNPFSSKGFALQFVSST 1140
 Db 1081 VHDVLRITWDGVEBEGVLLKELSGPALPKDLHSTFNSVVLQSTNPFSSKGFALQFVSST 1140
 QY 1141 ATSCNDGIGIPONGSRSDGMEAGDSVTFQCDPGYALQGSABEISCVKLENNRPFMOPSPCTC 1200
 Db 1141 ATSCNDGIGIPONGSRSDGMEAGDSVTFQCDPGYALQGSABEISCVKLENNRPFMOPSPCTC 1200
 QY 1201 IAPCGDLTGBSGVILSPNYPBPYPGKECDWKVTASPDVIALVFNIPULEPGYDFLHI 1260
 Db 1201 IAPCGDLTGBSGVILSPNYPBPYPGKECDWKVTASPDVIALVFNIPULEPGYDFLHI 1260
 QY 1261 YDGDLSJSLIGSYFGSOLPGRITSSNSLFLAFRSDASVSNAPFVLDIYENPRESFDP 1320
 Db 1261 YDGDLSJSLIGSYFGSOLPGRITSSNSLFLAFRSDASVSNAPFVLDIYENPRESFDP 1320
 QY 1321 GSINRGRVSGDLKLGSSTVYCHGGEVEGTSTLSCILGPDGPRVNNNRPVCTAACGG 1380
 Db 1321 GSINRGRVSGDLKLGSSTVYCHGGEVEGTSTLSCILGPDGPRVNNNRPVCTAACGG 1380
 QY 1381 QYVSGDGVLSPNYPONTSGOICLYVTVBKDYVFGQAFHFTALNDVVEVHDGSHOH 1440
 Db 1381 QYVSGDGVLSPNYPONTSGOICLYVTVBKDYVFGQAFHFTALNDVVEVHDGSHOH 1440

QY 1441 SRLISSSGHSGTESLPLATSNQVLIKFSAGKGLAPARGFFHYQAVPRTSATOCSSVPER 1500
 Db 1441 SRLISSSGHSGTESLPLATSNQVLIKFSAGKGLAPARGFFHYQAVPRTSATOCSSVPER 1500
 QY 1501 RYKRLGSDFSVGAIVRECNISGVALQGSBEIECLVPGLAOMVNSAPTCCVPCGGNLT 1560
 Db 1501 RYKRLGSDFSVGAIVRECNISGVALQGSBEIECLVPGLAOMVNSAPTCCVPCGGNLT 1560
 QY 1561 ERGCTILSPGPRPYNSLNCWKIIVREAGAIQIOVNSVTEOMNDSLEVPDGDNTVT 1620
 Db 1561 ERGCTILSPGPRPYNSLNCWKIIVREAGAIQIOVNSVTEOMNDSLEVPDGDNTVT 1620
 QY 1621 MGSFSGTVPALLNSTSNQLYLHFSYSDISVSAAGFHEXKTVGLSSCEPAPVSNQVKT 1680
 Db 1621 MGSFSGTVPALLNSTSNQLYLHFSYSDISVSAAGFHEXKTVGLSSCEPAPVSNQVKT 1680
 QY 1681 GERVLVNDVVSFOCEPQYALQGHANHSCEMGTVRMNYPPPLCTAOCGCTVEEMEGVILS 1740
 Db 1681 GERVLVNDVVSFOCEPQYALQGHANHSCEMGTVRMNYPPPLCTAOCGCTVEEMEGVILS 1740
 QY 1741 PGFPGNVPNSMDCSWKIALPVGFGAHQFANSTEPNHDYIEIRNGYETSRMNGRPSGS 1800
 Db 1741 PGFPGNVPNSMDCSWKIALPVGFGAHQFANSTEPNHDYIEIRNGYETSRMNGRPSGS 1800
 QY 1801 ELPSLSLSTSHETTVYFHSDSQNRPGFKLEYAYELOECRDEPFANGIVRAGAVNVQ 1860
 Db 1801 ELPSLSLSTSHETTVYFHSDSQNRPGFKLEYAYELOECRDEPFANGIVRAGAVNVQ 1860
 QY 1861 SVTFECLPGYQLTGHPLVLTQHGSTRNWDHPLRKECEVPCGGNITSSNGTVSFGPSPYS 1920
 Db 1861 SVTFECLPGYQLTGHPLVLTQHGSTRNWDHPLRKECEVPCGGNITSSNGTVSFGPSPYS 1920
 QY 1921 SSQDCWMLIVPFGHGRALMLSLLOTERPSGSDFTINDPGQTPRLCVFTRSMAKKTVO 1980
 Db 1921 SSQDCWMLIVPFGHGRALMLSLLOTERPSGSDFTINDPGQTPRLCVFTRSMAKKTVO 1980
 QY 1981 SSNOVLKFKPRDAGTGIFAIARSAVPLTCCPPTLIPNAEVTEENEFNIGDIVRYRCL 2040
 Db 1981 SSNOVLKFKPRDAGTGIFAIARSAVPLTCCPPTLIPNAEVTEENEFNIGDIVRYRCL 2040
 QY 2041 PGFTLVGNELITCKLGTLYLQFEGPPICEVHCPTEBLTJSTGVILSSQSYGSGYPOFQTC 2100
 Db 2041 PGFTLVGNELITCKLGTLYLQFEGPPICEVHCPTEBLTJSTGVILSSQSYGSGYPOFQTC 2100
 QY 2101 SWLVREPDVNIISLTYEYFLSEKQYDEFEIPDPSGQSLPKALSGVYSAPLIVTSSNS 2160
 Db 2101 SWLVREPDVNIISLTYEYFLSEKQYDEFEIPDPSGQSLPKALSGVYSAPLIVTSSNS 2160
 QY 2161 VYLRMSDHAHYNRKGFIRYSAPYCSLPRALHGFILGOTSTPGGSIFHGCNAGYLYG 2220
 Db 2161 VYLRMSDHAHYNRKGFIRYSAPYCSLPRALHGFILGOTSTPGGSIFHGCNAGYLYG 2220
 QY 2221 HSNALCTRHPOGYHLSSEALPLCOALSCLGPEAPKNGMVEGKEYVTGTRKAVYCSBEGYL 2280
 Db 2221 HSNALCTRHPOGYHLSSEALPLCOALSCLGPEAPKNGMVEGKEYVTGTRKAVYCSBEGYL 2280
 QY 2281 QAGABATACLDLGLMSNRNVPPOCVRESGNGSGVSTCDDVSIISVHERMRLIFETQY 2340
 Db 2281 QAGABATACLDLGLMSNRNVPPOCVRESGNGSGVSTCDDVSIISVHERMRLIFETQY 2340
 QY 2332 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTTCRIISCGELPIPNHGRIGTLS 2391
 Db 2332 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTTCRIISCGELPIPNHGRIGTLS 2391
 QY 2341 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTTCRIISCGELPIPNHGRIGTLS 2400
 Db 2341 QFOAOLMLICDPPGYTTCGVIRICQANGKMSLDSPTTCRIISCGELPIPNHGRIGTLS 2400
 QY 2392 VYGATAIIFS CNSGYTLVGRVRECMANGLWSGSEVRCL 2429
 Db 2392 VYGATAIIFS CNSGYTLVGRVRECMANGLWSGSEVRCL 2429
 QY 2401 VYGATAIIFS CNSGYTLVGRVRECMANGLWSGSEVRCL 2460
 Db 2401 VYGATAIIFS CNSGYTLVGRVRECMANGLWSGSEVRCL 2460
 QY 2430 ---AGHGTBEPVINGHINENSYRSGSVYVQCNAGPRLIGSVYRICQODHMSGKTPFC 2486
 Db 2430 ---AGHGTBEPVINGHINENSYRSGSVYVQCNAGPRLIGSVYRICQODHMSGKTPFC 2486
 QY 2461 LTRAGHGTBEPVINGHINENSYRSGSVYVQCNAGPRLIGSVYRICQODHMSGKTPFC 2520
 Db 2461 LTRAGHGTBEPVINGHINENSYRSGSVYVQCNAGPRLIGSVYRICQODHMSGKTPFC 2520
 QY 2487 V-----PITCGHGNPYNGLTQGNQFNLNDVVFVGNPGY 2521

DB 2521 VHVKKOQLLLLLLLLLLGGDDDDDDSGAITCGHPGVPVGLGQGNFNLNDVVKVVCNPGY 2560
 QY 2522 MAEGARSOCLASGQSMPLPTCRILINCTDGHQNSYRQVHAGPHRFSGTTSYRCN 2561
 DB 2561 MAEGARSOCLASGQSMPLPTCRILINCTDGHQNSYRQVHAGPHRFSGTTSYRCN 2640
 QY 2582 HGFYLLGTPVLSCQDGTMDRPPQCL 2608
 DB 2641 HGFYLLGTPVLSCQDGTMDRPPQCL 2667
 RESULT 9
 ID ADH71140 standard; protein; 2669 AA.
 AC ADH71140;
 DT 25-MAR-2004 (first entry)
 DE Human protein of the invention NOV4c SEQ ID NO:36.
 XX
 KW human; cytosolic; immunomodulator; neuroprotective; noctropic;
 KW anorectic; antidiabetic; antimicrobial; antihypaemic; gene therapy;
 KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
 KW obesity; diabetes; infectious disease; metabolic syndrome X;
 KW dyslipidaemia.
 XX
 OS Homo sapiens.
 XX WO000102155-A2.
 PD 11-DEC-2003.
 XX
 PF 03-JUN-2003; 2003WO-US017430.
 XX
 PR 03-JUN-2002; 2002US-0385120P.
 PR 04-JUN-2002; 2002US-0385784P.
 PR 05-JUN-2002; 2002US-0386041P.
 PR 06-JUN-2002; 2002US-0386047P.
 PR 06-JUN-2002; 2002US-0386376P.
 PR 06-JUN-2002; 2002US-0386453P.
 PR 06-JUN-2002; 2002US-0386864P.
 PR 06-JUN-2002; 2002US-0387016P.
 PR 07-JUN-2002; 2002US-038796P.
 PR 07-JUN-2002; 2002US-0386816P.
 PR 07-JUN-2002; 2002US-0386931P.
 PR 07-JUN-2002; 2002US-0386942P.
 PR 07-JUN-2002; 2002US-0386971P.
 PR 08-JUN-2002; 2002US-0387622P.
 PR 08-JUN-2002; 2002US-0296360P.
 PR 10-JUN-2002; 2002US-0387400P.
 PR 10-JUN-2002; 2002US-0387535P.
 PR 11-JUN-2002; 2002US-0387610P.
 PR 11-JUN-2002; 2002US-0387625P.
 PR 11-JUN-2002; 2002US-0387634P.
 PR 11-JUN-2002; 2002US-0387668P.
 PR 11-JUN-2002; 2002US-0387696P.
 PR 11-JUN-2002; 2002US-0387702P.
 PR 11-JUN-2002; 2002US-0387836P.
 PR 11-JUN-2002; 2002US-0387859P.
 PR 12-JUN-2002; 2002US-0387933P.
 PR 12-JUN-2002; 2002US-0387934P.
 PR 12-JUN-2002; 2002US-0387960P.
 PR 12-JUN-2002; 2002US-0388022P.
 PR 12-JUN-2002; 2002US-0388022P.
 PR 13-JUN-2002; 2002US-0388096P.
 PR 14-JUN-2002; 2002US-0389123P.
 PR 14-JUN-2002; 2002US-0389118P.
 PR 14-JUN-2002; 2002US-0389120P.
 PR 14-JUN-2002; 2002US-0389144P.
 PR 14-JUN-2002; 2002US-0389146P.
 PR 17-JUN-2002; 2002US-0389729P.
 PR 17-JUN-2002; 2002US-0389742P.

PR 18-JUN-2002; 2002US-0389884P.
 PR 19-JUN-2002; 2002US-039006P.
 PR 19-JUN-2002; 2002US-0390203P.
 PR 21-JUN-2002; 2002US-0390763P.
 PR 17-JUL-2002; 2002US-0396706P.
 PR 06-AUG-2002; 2002US-0401628P.
 PR 09-AUG-2002; 2002US-0402156P.
 PR 09-AUG-2002; 2002US-040256P.
 PR 09-AUG-2002; 2002US-040289P.
 PR 12-AUG-2002; 2002US-0402786P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 13-AUG-2002; 2002US-0402832P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403531P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 13-AUG-2002; 2002US-0406317P.
 PR 15-AUG-2002; 2002US-0406317P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0410084P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-0412731P.
 PR 30-SEP-2002; 2002US-0414801P.
 PR 30-SEP-2002; 2002US-0414839P.
 PR 30-SEP-2002; 2002US-0414840P.
 PR 30-SEP-2002; 2002US-0414954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.
 PA (CURA)-CURAGEN CORP.
 XX
 PI Alsobrook JP, Alvarez E, Anderson DW, Boldog FL, Casman SJ;
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
 PI Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
 PI Guesey VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, Macdougall JR;
 PI MacLachlan T, Malyankar UM, Mezick AJ, Millet I, Mishra VS;
 PI Padigaru W, Paturajan M, Pena CE, Peyman JA, Raha D, Rastelli L;
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkets RA;
 PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
 PI Zhong H;
 XX
 DR WPI; 2004-081935/08.
 DR N-PSDB; ADH71139.
 XX
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 PR
 SS Example 4; SEQ ID NO 36; 1880pp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (NOVX). A
 CC polypeptide of the invention has cytosolic, immunomodulator,
 CC neuroprotective, noctropic, anorectic, antidiabetic, antimicrobial, and
 CC antihypaemic activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing, the
 CC treating or preventing NOVX-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemia. The nucleic acids are

CC further used as hybridisation probes, in chromosome mapping, tissue
CC typing, preventive medicine, and pharmacogenetics. The present sequence
CC represents a NOXV polypeptide of the invention.

Sequence 2669 AA:

Query Match 83.3%; Score 14142.5; DB 8; Length 2669;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 2605; Conservative 1; Mismatches 2; Indels 59; Gaps 3;

QY 1 MAGAPPALLPCLSDCCASQNRHSVGVSELVKKQIELKSGVKKMPSKXQKTS 60
DB 1 MAGAPPALLPCLSDCCASQNRHSVGVSELVKKQIELKSGVKKMPSKXQKTS 60
QY 61 VLTQVGSQGHNMCPDGGIPERGRKLSDFRLGSSVFTCEGVDLOGSKRITCMKSDM 120
DB 61 VLTQVGSQGHNMCPDGGIPERGRKLSDFRLGSSVFTCEGVDLOGSKRITCMKSDM 120
QY 121 FAAMSDRPPVCRAACAHLRGPGIITSPNFIQYNNNAHCWIIITALNPSKVIKLAPE 180
DB 121 FAAMSDRPPVCRAACAHLRGPGIITSPNFIQYNNNAHCWIIITALNPSKVIKLAPE 180
QY 181 EFDLEERYDTLVYDGGQDQDKTVLYMSQVACSDSPHTPSRIPEMSGDIMRQKXTVL 240
DB 181 EFDLEERYDTLVYDGGQDQDKTVLYMSQVACSDSPHTPSRIPEMSGDIMRQKXTVL 240
QY 241 EICRDISSDARSVSKSPKTSNAVELVAPGTEIEGSGCDPPIPAYGRREGRFHGD 300
DB 241 EICRDISSDARSVSKSPKTSNAVELVAPGTEIEGSGCDPPIPAYGRREGRFHGD 300
QY 301 TLKECPAPFELVGOKAITCOKNQMSAKKPGCVFCFNPFTSPSGVILSPNYPEDYGNH 360
DB 301 TLKECPAPFELVGOKAITCOKNQMSAKKPGCVFCFNPFTSPSGVILSPNYPEDYGNH 360
QY 361 LHCWMLIABPESRIHLAFNDIDVEPODFLVIKDGATABAPVLTGFSGNQLPSSITSG 420
DB 361 LHCWMLIABPESRIHLAFNDIDVEPODFLVIKDGATABAPVLTGFSGNQLPSSITSG 420
QY 421 HVARLEQTHSTKRGFNITFTFRNECPDPGVPNVGRKFGDSLQSGSISFLCEBGF 480
DB 421 HVARLEQTHSTKRGFNITFTFRNECPDPGVPNVGRKFGDSLQSGSISFLCEBGF 480
QY 481 LGTQSGEITTCVLEKGSVWNSAVLRCBAPCGHLTSPSGTILSPGMGFKDALSCAW 540
DB 481 LGTQSGEITTCVLEKGSVWNSAVLRCBAPCGHLTSPSGTILSPGMGFKDALSCAW 540
QY 541 IEAOPGYPIKITPDRFTEVNYDTLEVRDGRYSAPLIGVYHGTQVQFLISTSNLYL 600
DB 541 IEAOPGYPIKITPDRFTEVNYDTLEVRDGRYSAPLIGVYHGTQVQFLISTSNLYL 600
QY 601 PSTDKSHSDIGFQRLRYETITLQSDHCLDPGIPVNGQHRGNDPYGALVTFSGSGYTLSD 660
DB 601 PSTDKSHSDIGFQRLRYETITLQSDHCLDPGIPVNGQHRGNDPYGALVTFSGSGYTLSD 660
QY 661 GBPLECPNFMQMSALPSCALCGGFIQSSGTLSPGPPFVNNNLCMIIETSHSGK 720
DB 661 GBPLECPNFMQMSALPSCALCGGFIQSSGTLSPGPPFVNNNLCMIIETSHSGK 720
QY 721 VFTFHTFHLSESHDYLITENGSTFQPLQRLQTLGSRPAPISAGLYNFTAOVRFISDFS 780
DB 721 VFTFHTFHLSESHDYLITENGSTFQPLQRLQTLGSRPAPISAGLYNFTAOVRFISDFS 780
QY 781 MSYEGFHTTSEYVLEPCBEPEVAYISIRKGLQGVADTILFSCFPYRLEGTRITCLG 840
DB 781 MSYEGFHTTSEYVLEPCBEPEVAYISIRKGLQGVADTILFSCFPYRLEGTRITCLG 840
QY 841 GRRRLMSPLPRCAEAGNSVTGQTLSPNPNVNNNNHCYSLQTOBKGKIQOLAR 900
DB 841 GRRRLMSPLPRCAEAGNSVTGQTLSPNPNVNNNNHCYSLQTOBKGKIQOLAR 900
QY 901 APFELSEGDVLYKVDGNNNSARLLGVFSHSEMGVTLNSTSSLWLDITDAENTSKGPEL 960
DB 901 APFELSEGDVLYKVDGNNNSARLLGVFSHSEMGVTLNSTSSLWLDITDAENTSKGPEL 960

QY 961 HFSSFEIKCEDDGTPEKGYKVHDEGHFAGSSVSFSCDPPYSLRGSEILCLSGERRTWD 1020
DB 961 HFSSFEIKCEDDGTPEKGYKVHDEGHFAGSSVSFSCDPPYSLRGSEILCLSGERRTWD 1020
QY 1021 RPLPTCAECCGCTVRGEVSGQVLSPGYPAYEHNLCIMTIEAAGCTIGLHFLVPTTEE 1080
DB 1021 RPLPTCAECCGCTVRGEVSGQVLSPGYPAYEHNLCIMTIEAAGCTIGLHFLVPTTEE 1080
QY 1081 VHDVLRIWDGPVSEGVILKELSGPALPKDILHSTFNSVTLQSTDFPFSKOGFAIQESVST 1140
DB 1081 VHDVLRIWDGPVSEGVILKELSGPALPKDILHSTFNSVTLQSTDFPFSKOGFAIQESVST 1140
QY 1141 ATSCNDPICIPONSRSGSDWEAGDSTVFCQDPGALQSAEISCVKLENFFMOPSPPTC 1200
DB 1141 ATSCNDPICIPONSRSGSDWEAGDSTVFCQDPGALQSAEISCVKLENFFMOPSPPTC 1200
QY 1201 IAPCGDLTGPSCGVIILSPNPEEYPPGKECDMKVTYSPDVIALVFNIFMLEPGYDPLHI 1260
DB 1201 IAPCGDLTGPSCGVIILSPNPEEYPPGKECDMKVTYSPDVIALVFNIFMLEPGYDPLHI 1260
QY 1261 YDGRDLSPLIGSFYGSQLPGRIESSNSLFLAFRSDASVSNAGFVIDYENPRESCFDP 1320
DB 1261 YDGRDLSPLIGSFYGSQLPGRIESSNSLFLAFRSDASVSNAGFVIDYENPRESCFDP 1320
QY 1321 GSINKGRVSGDLKLGSSVTVYCHGGYEVGEGTSLSCILPDDGPPVNNNRPVCTAPCGG 1380
DB 1321 GSINKGRVSGDLKLGSSVTVYCHGGYEVGEGTSLSCILPDDGPPVNNNRPVCTAPCGG 1380
QY 1381 QYVSGDGVLSPNVYQNTYSGQICLYEVTVPKQYVVGQAPFHTALNDVVEVHDGSH 1440
DB 1381 QYVSGDGVLSPNVYQNTYSGQICLYEVTVPKQYVVGQAPFHTALNDVVEVHDGSH 1440
QY 1441 SRIILSLSGHTGESLPLATSNQVLIKFSAKGLAPARGFHVYQAVRTSATOCSSVPEP 1500
DB 1441 SRIILSLSGHTGESLPLATSNQVLIKFSAKGLAPARGFHVYQAVRTSATOCSSVPEP 1500
QY 1501 RYGRKLGSDPESVGAIVAFECNSGVALQGSPEICLPVGAALQMNVSAPTCVPPCGNLT 1560
DB 1501 RYGRKLGSDPESVGAIVAFECNSGVALQGSPEICLPVGAALQMNVSAPTCVPPCGNLT 1560
QY 1561 ERGRTILSPGFPEPYLSLNCWKIIVPEBAGIQIQVSVFTEQNMDSLEVPDQADTVT 1620
DB 1561 ERGRTILSPGFPEPYLSLNCWKIIVPEBAGIQIQVSVFTEQNMDSLEVPDQADTVT 1620
QY 1621 MUGSPSGTVPALLNSTSNQYLHFYSDISVSAAGFLEKTVGLSSCPREPAVPSNGVKT 1680
DB 1621 MUGSPSGTVPALLNSTSNQYLHFYSDISVSAAGFLEKTVGLSSCPREPAVPSNGVKT 1680
QY 1681 GERLYLVNDVVSFOCEPEYALQGHANISCMGTVARMYPPPLCIAQGGTVEEMEGVILS 1740
DB 1681 GERLYLVNDVVSFOCEPEYALQGHANISCMGTVARMYPPPLCIAQGGTVEEMEGVILS 1740
QY 1741 PGFPYVPSNMDCSWKIALPVFGAHIQFNFTSEPHVDYIIRNGPYETSRMMGRSGS 1800
DB 1741 PGFPYVPSNMDCSWKIALPVFGAHIQFNFTSEPHVDYIIRNGPYETSRMMGRSGS 1800
QY 1801 ELPSLSLSTSHETTVYFHSQNSQNRPGFKLEYAYELQECRDEPFANGIVRAGYVNGQ 1860
DB 1801 ELPSLSLSTSHETTVYFHSQNSQNRPGFKLEYAYELQECRDEPFANGIVRAGYVNGQ 1860
QY 1861 SVTFECLPGYQLGHPLVLTQHGTRNRWDBPLKCEVPCGGNITSSNGTYYSQFPSPYS 1920
DB 1861 SVTFECLPGYQLGHPLVLTQHGTRNRWDBPLKCEVPCGGNITSSNGTYYSQFPSPYS 1920
QY 1921 SSQDCWLVITVPIGHGRNLNLSLQTEPSGDFTTIMDGPQOTAPRLGVFTRSMAKKTQVS 1980
DB 1921 SSQDCWLVITVPIGHGRNLNLSLQTEPSGDFTTIMDGPQOTAPRLGVFTRSMAKKTQVS 1980
QY 1981 SSNQVLLKFRDAATGIGIFAIAFSAYPLTKCPPTILPNAEVVTENEEFNIGDI VRYRCL 2040
DB 1981 SSNQVLLKFRDAATGIGIFAIAFSAYPLTKCPPTILPNAEVVTENEEFNIGDI VRYRCL 2040

PA (CURA-) CURAGEN CORP.
 XX
 PI Alabrook JP, Alvarez E, Anderson DM, Boldog FI, Caeman SJ,
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K,
 PI Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X,
 PI Gusev VY, Hermann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR,
 PI MacLachlan T, Malpankar UM, Mexick AJ, Millet I, Mishra VS;
 PI Padiganu M, Patutarjan M, Pena CE, Peyman JA, Raha D, Rastelli L,
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shimkets RA,
 PI Smutson G, Spytek KA, Stone DU, Vernet CM, Voss EZ, Zhong M,
 PI Zhong H;
 DR WPI: 2004-081935/08.
 DR N-PSDB: ADH71141.
 PT New NOVX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOVX-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
 XX
 PS Example 4; SEQ ID NO 38; 1880bp; English.
 XX
 CC The invention relates to a novel isolated polypeptide (NOVX). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antiparasitic activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOVX polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosis,
 CC treatment or preventing NOVX-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, obesity, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acids are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOVX polypeptide of the invention.
 XX
 SQ Sequence 2612 AA:
 Query Match 77.4%; Score 13142.5; DB 8; Length 2612;
 Best Local Similarity 91.6%; Pred. No. 0;
 Matches 2457; Conservative 20; Mismatches 57; Indels 147; Gaps 12;

Db 421 HVARLEFQTDHSTGKRGFNITFTFRHNECPDGPVNVGKRFGLSLQAGSSISFLCDEGF 480
 QY 481 LGQGSFTITCVLKEGVSVMNSAVLRCEACGGHLSPSSTILSPGKRGFKALSCAMV 540
 Db 481 LGQGSFTITCVLKEGVSVMNSAVLRCEACGGHLSPSSTILSPGKRGFKALSCAMV 540
 QY 541 IEAOPGPIKITDRFTEVNVYTLERDRRTYSAPLIGVYHQVQPOFLISTSNLYL 600
 Db 541 IEAOPGPIKITDRFTEVNVYTLERDRRTYSAPLIGVYHQVQPOFLISTSNLYL 600
 QY 601 FSTDKSHSDIGFQLYREYITTLQSDHCLDPGIPVNGQHGNDFYVALVTFSCSGYTLSD 660
 Db 601 FSTDKSHSDIGFQLYREYITTLQSDHCLDPGIPVNGQHGNDFYVALVTFSCSGYTLSD 660
 QY 661 GEPLCEPFPNQRALPSCALCGGFIQSSGSTITLSPGFPDPNNLNCMIETSHGK 720
 Db 661 GEPLCEPFPNQRALPSCALCGGFIQSSGSTITLSPGFPDPNNLNCMIETSHGK 720
 QY 721 VFPTFTFHLESQHDYLLITENGSPFOPLRLQSGRLPAPISAGLYGNFTAQRFIIDS 780
 Db 721 VFPTFTFHLESQHDYLLITENGSPFOPLRLQSGRLPAPISAGLYGNFTAQRFIIDS 780
 QY 781 MSYEGFNITSEYDLBCEBEPVAYSIRKGLQGVGDTLTFSCFPGYRLEGTAIRITCLG 840
 Db 781 MSYEGFNITSEYDLBCEBEPVAYSIRKGLQGVGDTLTFSCFPGYRLEGTAIRITCLG 840
 QY 841 GRRRLWSSPLPRCAVEGNSVLTGTQTLSPNFPVNNNHECTYSIQTOPGKIQOKAR 900
 Db 841 GRRRLWSSPLPRCAVEGNSVLTGTQTLSPNFPVNNNHECTYSIQTOPGKIQOKAR 900
 QY 901 AFELSEBDYLVKVDGNNSARLLGVPSHSMGVTLNSTSSLMLDITAEYNTSKFEL 960
 Db 901 AFELSEBDYLVKVDGNNSARLLGVPSHSMGVTLNSTSSLMLDITAEYNTSKFEL 960
 QY 961 HFSSFELICEKEDPTKFGYKVAHDEGHFAGSVSFSCDPGYSLRGSEELLCSGERRTWD 1020
 Db 961 HFSSFELICEKEDPTKFGYKVAHDEGHFAGSVSFSCDPGYSLRGSEELLCSGERRTWD 1020
 QY 1021 RPLPTCAVECGTVRGEVSGQVTSFGYPAHYENHNCIWTIEBAGCTTGLHLVPTTEE 1080
 Db 1021 RPLPTCAVECGTVRGEVSGQVTSFGYPAHYENHNCIWTIEBAGCTTGLHLVPTTEE 1080
 QY 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDLSHFNSVVLQSFDPFYSKQPAIQFVST 1140
 Db 1081 VHDVLRIMDGPVBSGVLLKELSGPALPKDLSHFNSVVLQSFDPFYSKQPAIQFVST 1140
 QY 1141 ATSCNDPGIPQNSRSRSDSWEAGDSTVFQCDPGYALQGSALIESCVKIEENRFPMQSPPTC 1200
 Db 1141 ATSCNDPGIPQNSRSRSDSWEAGDSTVFQCDPGYALQGSALIESCVKIEENRFPMQSPPTC 1200
 QY 1201 IAPCGDLPSPGVILSPNYPEPYPGKECDKVTYVSDYVIALVFNIPNLEPGYPLHI 1260
 Db 1201 IAPCGDLPSPGVILSPNYPEPYPGKECDKVTYVSDYVIALVFNIPNLEPGYPLHI 1260
 QY 1261 YDGRDSLPLIGSFYSGQLPGRLESSNSLFLFRSPASVSNAGFVLDYENPRESCEFD 1320
 Db 1261 YDGRDSLPLIGSFYSGQLPGRLESSNSLFLFRSPASVSNAGFVLDYENPRESCEFD 1320
 QY 1321 GSINKGTRVSGDLKLGSSVYYCHGGEVGTSLSCILGPDGKPVNNNRPVCTAPCGG 1380
 Db 1321 GSINKGTRVSGDLKLGSSVYYCHGGEVGTSLSCILGPDGKPVNNNRPVCTAPCGG 1380
 QY 1381 QYVSGDGVILSPNYPONTYSGQICLVFVTVPKDYVVFQGFAPFHTALNDVVEYHDSQH 1440
 Db 1381 QYVSGDGVILSPNYPONTYSGQICLVFVTVPKDYVVFQGFAPFHTALNDVVEYHDSQH 1440
 QY 1441 SRLLSLSGSHTESLPLATSNQVLIKFSKGLAPARGFFVYQAVPRSATQCSSVPEP 1500
 Db 1441 SRLLSLSGSHTESLPLATSNQVLIKFSKGLAPARGFFVYQAVPRSATQCSSVPEP 1500
 QY 1501 RYGRRLGSDPSVGAIVRFECSNGYALQGSPEIECLPVGALQAMNVSAPTCVPCGGLT 1560
 Db 1501 RYGRRLGSDPSVGAIVRFECSNGYALQGSPEIECLPVGALQAMNVSAPTCVPCGGLT 1560

QY 1561 ERRGIIISPGPEPIYNSLNCWVKIVPEGAGIOIOVVSFVTEQNMWDSLEVPDGDANTVT 1620
DB 1561 ERRTIISPGPEPIYNSLNCWVKIVPEGAGIOIOVVSFVTEQNMWDSLEVPDGDANTVT 1620
QY 1621 MLGSESGTIVALLNSTNSQYLHPYSDISVSAAGPHLEKTVGLSSCPREPAVPSNGVKT 1680
DB 1621 MLGSESGTIVALLNSTNSQYLHPYSDISVSAAGPHLEKTVGLSSCPREPAVPSNGVKT 1680
QY 1681 GERLYVNDVVSFOCEPGYALOGHAHISCMPTVARNVYPPRLCTAOCGGVTEEMEGYILS 1740
DB 1681 GERLYVNDVVSFOCEPGYALOGHAHISCMPTVARNVYPPRLCTAOCGGVTEEMEGYILS 1740
QY 1741 PGFPGNYPNSMNDCSWKIALPVFGAHIQPLNFSTERNHDYIEIRNGEYETSRMMGRPSGS 1800
DB 1741 PGFPGNYPNSMNDCSWKIALPVFGAHIQPLNFSTERNHDYIEIRNGEYETSRMMGRPSGS 1800
QY 1801 ELPSLSLSTSHETTVYHSHDSQNRPGFKLEYQAYELQECRDPPEPPANGIVRAGYVVGQ 1860
DB 1801 ELPSLSLSTSHETTVYHSHDSQNRPGFKLEYQAYELQECRDPPEPPANGIVRAGYVVGQ 1860
QY 1861 SVTEFCLPGYOLTGHPVLTGCOHGNRMNDHPLPKCEVCGGNTSSNGTVYSPGSPSPYS 1920
DB 1861 SVTEFCLPGYOLTGHPVLTGCOHGNRMNDHPLPKCEVCGGNTSSNGTVYSPGSPSPYS 1920
QY 1921 SSODCVWLITVPIGHGVRLNLSLQTEPSGDFITIMDGPQOTAPRLGVFTSRMAKTVQS 1980
DB 1921 SSODCVWLITVPIGHGVRLNLSLQTEPSGDFITIMDGPQOTAPRLGVFTSRMAKTVQS 1980
QY 1981 SSNOVLKPHRDATAAGIIPAIASAVPLTKCPPTLIPNAEVTNENEFNIGIVYRCL 2040
DB 1981 SSNOVLKPHRDATAAGIIPAIASAVPLTKCPPTLIPNAEVTNENEFNIGIVYRCL 2040
QY 2041 PGFTLVGNELITCKLGTYLOFEGEPPICEVHCPTNELITDSTGVILSOSYPSGSPQOTC 2100
DB 2041 PGFTLVGNELITCKLGTYLOFEGEPPICEVHCPTNELITDSTGVILSOSYPSGSPQOTC 2100
QY 2101 SWLVREPDYINISITVEYFLSEKQYDFEIFDPGSGSPILKALSGNYSAPLIVTSSNS 2160
DB 2101 SWLVREPDYINISITVEYFLSEKQYDFEIFDPGSGSPILKALSGNYSAPLIVTSSNS 2160
QY 2161 VYLRMSDHAYNRKGFIIRASAPYCSLPRAPLHGFILGOSTPGSIHFCNAGRLVG 2220
DB 2161 VYLRMSDHAYNRKGFIIRASAPYCSLPRAPLHGFILGOSTPGSIHFCNAGRLVG 2220
QY 2221 HSMALCTHRDQYHLSSEALPLCOALSCGLPEAPKNGVFGKEYTVGTAVYSCSEGYHL 2280
DB 2221 HSMALCTHRDQYHLSSEALPLCOALSCGLPEAPKNGVFGKEYTVGTAVYSCSEGYHL 2280
QY 2281 QAGAEATAECTDGTGLMSNRNVPPQCVPTCPDVYSISVHGRWRLIFETQYFOAQLMLI 2340
DB 2281 QAGAEATAECTDGTGLMSNRNVPPQCVPTCPDVYSISVHGRWRLIFETQYFOAQLMLI 2340
QY 2341 CDPGYVYTGQVRVRCQANGKMSLGDSTPCRIISCEGLPIPNHRIIGLTVYGAATAFS 2400
DB 2341 CDPGYVYTGQVRVRCQANGKMSLGDSTPCRIISCEGLPIPNHRIIGLTVYGAATAFS 2400
QY 2401 CUSGYTVLSRVRECMANGLSGSEVRCLAGHGTPEPIVNGHGENYSYRSGVYVQCN 2460
DB 2401 CUSGYTVLSRVRECMANGLSGSEVRCLAGHGTPEPIVNGHGENYSYRSGVYVQCN 2460
QY 2461 AGFRLIGMSVRIQQODHHSKGTBPC-VPIITGCHRNPNVNGLTGQNFRLNDVVKVCNP 2519
DB 2461 AGFRLIGMSVRIQQODHHSKGTBPC-VPIITGCHRNPNVNGLTGQNFRLNDVVKVCNP 2519
QY 2519 AGFRLIGMSVRIQQODHHSKGTBPC-VPIITGCHRNPNVNGLTGQNFRLNDVVKVCNP 2519
DB 2519 AGFRLIGMSVRIQQODHHSKGTBPC-VPIITGCHRNPNVNGLTGQNFRLNDVVKVCNP 2519
QY 2550 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562
DB 2550 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562
QY 2562 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562
DB 2562 GYMAEGARSOCLASQMSDMLPTCRI-----INCTDPGHQENSVRQV- 2562
QY 2595 DTKKIVFVASTPAARTGSASAPLCLTGATTASTSWAPQCSARBAWAG--TVPAVSQVS 2595
DB 2595 DTKKIVFVASTPAARTGSASAPLCLTGATTASTSWAPQCSARBAWAG--TVPAVSQVS 2595

QY 2596 GDSWTDRPRPQCL-----LVSCGHRSPPHSQMSDSTYVG 2631
DB 2547 SGVLPWPSGLPASPDPVWRQLYCGSSGA---VQLHROAVSAG 2584
RESULT 11
ADH72216
ID ADH72216 standard; protein; 3567 AA.
XX
AC ADH72216;
XX
DT 25-MAR-2004 (first entry)
XX
DE Human protein of the invention NOV54b SEQ ID NO:1112.
XX
KW human; cytostatic; immunomodulator; neuroprotective; nootropic;
KW anorectic; antidiabetic; antimicrobial; antidiabetic; gene therapy;
KW vaccine; cancer; cachexia; Alzheimer's disease; Parkinson's disease;
KW obesity; diabetes; infectious disease; metabolic syndrome X;
KW dyslipidaemia.
XX
OS Homo sapiens.
XX
PN WO2003102155-A2.
XX
PD 11-DEC-2003.
XX
PF 03-JUN-2003; 2003WO-US017430.
XX
XX
03-JUN-2002; 2002US-0385120P.
PR 04-JUN-2002; 2002US-0385784P.
PR 05-JUN-2002; 2002US-0386041P.
PR 05-JUN-2002; 2002US-0386047P.
PR 06-JUN-2002; 2002US-0386376P.
PR 06-JUN-2002; 2002US-0386435P.
PR 06-JUN-2002; 2002US-0386864P.
PR 06-JUN-2002; 2002US-0387016P.
PR 07-JUN-2002; 2002US-0386796P.
PR 07-JUN-2002; 2002US-0386816P.
PR 07-JUN-2002; 2002US-0386931P.
PR 07-JUN-2002; 2002US-0386942P.
PR 07-JUN-2002; 2002US-0386971P.
PR 07-JUN-2002; 2002US-0387262P.
PR 08-JUN-2002; 2002US-0296960P.
PR 10-JUN-2002; 2002US-0387400P.
PR 10-JUN-2002; 2002US-0387535P.
PR 11-JUN-2002; 2002US-0387610P.
PR 11-JUN-2002; 2002US-0387625P.
PR 11-JUN-2002; 2002US-0387634P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387668P.
PR 11-JUN-2002; 2002US-0387702P.
PR 11-JUN-2002; 2002US-0387836P.
PR 11-JUN-2002; 2002US-0387859P.
PR 12-JUN-2002; 2002US-0387933P.
PR 12-JUN-2002; 2002US-0387934P.
PR 12-JUN-2002; 2002US-0387960P.
PR 12-JUN-2002; 2002US-0388022P.
PR 12-JUN-2002; 2002US-0388096P.
PR 13-JUN-2002; 2002US-0389123P.
PR 14-JUN-2002; 2002US-0389118P.
PR 14-JUN-2002; 2002US-0389120P.
PR 14-JUN-2002; 2002US-0389144P.
PR 14-JUN-2002; 2002US-0389146P.
PR 14-JUN-2002; 2002US-0389729P.
PR 17-JUN-2002; 2002US-0389742P.
PR 18-JUN-2002; 2002US-0389884P.
PR 19-JUN-2002; 2002US-0390060P.
PR 19-JUN-2002; 2002US-0390209P.
PR 21-JUN-2002; 2002US-0390763P.
PR 17-JUL-2002; 2002US-0396706P.
PR 06-AUG-2002; 2002US-0401628P.
PR 09-AUG-2002; 2002US-0402156P.

PR 09-AUG-2002; 2002US-0402256P.
 PR 09-AUG-2002; 2002US-0402389P.
 PR 12-AUG-2002; 2002US-0402786P.
 PR 12-AUG-2002; 2002US-0402816P.
 PR 12-AUG-2002; 2002US-0402821P.
 PR 12-AUG-2002; 2002US-0402832P.
 PR 13-AUG-2002; 2002US-0403448P.
 PR 13-AUG-2002; 2002US-0403459P.
 PR 13-AUG-2002; 2002US-0403511P.
 PR 13-AUG-2002; 2002US-0403532P.
 PR 13-AUG-2002; 2002US-0403563P.
 PR 13-AUG-2002; 2002US-0406317P.
 PR 15-AUG-2002; 2002US-0406317P.
 PR 26-AUG-2002; 2002US-0406182P.
 PR 26-AUG-2002; 2002US-0406355P.
 PR 27-AUG-2002; 2002US-0406240P.
 PR 12-SEP-2002; 2002US-0412528P.
 PR 20-SEP-2002; 2002US-0412528P.
 PR 23-SEP-2002; 2002US-0412731P.
 PR 30-SEP-2002; 2002US-0414801P.
 PR 30-SEP-2002; 2002US-0414839P.
 PR 30-SEP-2002; 2002US-0414840P.
 PR 30-SEP-2002; 2002US-0414954P.
 PR 09-OCT-2002; 2002US-0417186P.
 PR 09-OCT-2002; 2002US-0417406P.
 PR 23-OCT-2002; 2002US-0420639P.
 PR 28-OCT-2002; 2002US-0421156P.
 PR 31-OCT-2002; 2002US-0422690P.
 PR 01-NOV-2002; 2002US-0423130P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 05-NOV-2002; 2002US-0423798P.
 PR 12-NOV-2002; 2002US-0425453P.

(CURA-) CURAGEN CORP.

PI Albrook JP, Alvarez E, Anderson DM, Boldog FE, Caaman SJ;
 PI Catterton E, Chapoval A, Crabtree-Bokor JR, Edinger SR, Ellerman K;
 PI Ettenberg S, Gangoli EA, Gerlach VL, Gorman L, Gunther E, Guo X;
 PI Gusev VY, Herrmann JL, Ji W, Kekuda R, Li L, Liu X, MacDougall JR;
 PI MacLachlan T, Malyanek UM, Mezick AJ, Millet I, Mishra VS;
 PI Padiguru M, Paturajan M, Pena CE, Peyman JA, Raha D, Raetelli L;
 PI Rieger DK, Rothenberg ME, Sciore P, Shenoy SG, Shinkels RA;
 PI Smithson G, Spytek KA, Stone DJ, Vernet CM, Voss EZ, Zhong M;
 PI Zhong H;

WPI; 2004-081935/08.
 DR N-PSDB; ADH72215.

PT New NOX polypeptides and nucleic acid molecules useful for preventing or
 PT treating NOX-associated disorders, e.g. cancer, diabetes, infection or
 PT obesity, and in chromosome mapping, tissue typing or pharmacogenomics.

XX Example 54; SEQ ID NO 112; 1880bp; English.

CC The invention relates to a novel isolated polypeptide (NOX). A
 CC polypeptide of the invention has cytostatic, immunomodulator,
 CC neuroprotective, nootropic, anorectic, antidiabetic, antimicrobial, and
 CC antilipase activity, and may have a use in gene therapy, and as a
 CC vaccine. The polypeptides are encoded by NOX polynucleotides comprising
 CC any of the 303 fully defined nucleotide sequences given in the
 CC specification. The polypeptide is useful in the manufacture of a
 CC medicament for treating a syndrome associated with a human disease. The
 CC polypeptide, polynucleotide and antibody are useful in diagnosing,
 CC treating or preventing NOX-associated disorders, e.g. cancer, cachexia,
 CC Alzheimer's disease, Parkinson's disease, diabetes, infectious
 CC diseases, metabolic syndrome X or dyslipidaemias. The nucleic acid are
 CC further used as hybridisation probes, in chromosome mapping, tissue
 CC typing, preventive medicine, and pharmacogenomics. The present sequence
 CC represents a NOX polypeptide of the invention.

XX Sequence 3567 AA;

Query Match 66.2%, Score 11247.5; DB 8; Length 3567;

Best Local Similarity 63.6%; Pred. No. 0;
 Matches 1957; Conservative 458; Mismatches 600; Indels 61; Gaps 11;
 QY 1 MAGAPPALI-----LPCSLISD-----CCASQNRHSVGVSELYKKQJELKSRGV 47
 Db 264 ISGTEAPSIWLTGMNLPSPVSSKNNMLRLHFTSDSNRRK-GPNAQFOYVKAILEKSRGV 322
 QY 48 KLMPSKXNSOKTSVLTQVGVSGAHNCCPDGIERGGRIGSDP-RIGSSVQFCTNEGYDL 106
 Db 323 KMLPSKXNSKNSVLSQGVVALVSHMCLDPIENGNGRAGSDSRVAGANVQFSCEDNYVL 382
 QY 107 QGSKRITCMKVSMDFAWSDHRRPVCRAHMDALRGPSGIITSPNPIQYNNAHCVIIL 166
 Db 383 QGSKSITCQVETETLAAMSHPRIQARTGSLNLRGSGVITSPNPIQYNNAHCVIIL 442
 QY 167 TALNPSKVIKLAFEEFDLERGYDTLVGDDGQDQKTVLYMSQACSDSPHTPSRIPE 226
 Db 443 TTIDPDKVIKLAFEEFDLERGYDTLVGDDGQDQKTVLYV-----LTGSVDP 492
 QY 227 ---SMGDIWROKMTVLEICRODISSDARSGSVKSPKTSNAYELVARG-----TEIEOG 278
 Db 493 LIVSMNSQMWLH-----LQSDS-----ISPGKRAYQIEK 526
 QY 279 SCGDPGIPAYGRREGSRFHFGDTLKFECPAPAFELVGOIKAITCOKNQWASAKKFCVSCF 338
 Db 527 GCGDPGIPAYGKRTGSSFLHGDITLFECPAPAFELVGERVITCOQNNQWASAKKFCVSCF 586
 QY 339 FNFTSPGVVLSBNYPEDYGNHLHCVLILAREBSRIHLAFNDIIVEPQDFLVIKDAT 398
 Db 587 FNFTASGIIISBNYPEDYGNHLHCVLILAREBSRIHLAFNDIIVEPQDFLVIKDAT 646
 QY 399 AEAIVLCTFSGNQLPSSITSSGHVAPLEPCTDSTGKRGNIITFTFRHNECDPGVNV 458
 Db 647 SDITVLTGTFSGNVPQLASSGHI VRLFEFSDHSTTGKRGNIITFTFRHNECDPGVNV 706
 QY 459 GKRFDSILQSSISIFLDCDGFAGTQSEITITVLKESGVVMSAVLRCAPCGHILSP 518
 Db 707 GKRFGDRFLGSSVSHFCDGFKVTKQSEBITCILQDGNVMSVTRCACPCHILAS 766
 QY 519 SGTLSPGWGFFKDALSCAMVIEAOPGYIKTTPDEKTEVNYDTLEVDGRTYSAPLI 578
 Db 767 SGVTLPGWGVYKXDSHCEWIEAKRGHSIKMTPDFQEVNVYDTLEVDGRTYSAPLI 826
 QY 579 GYHGTQVPOPLISTSVYLLFTSTKSHSDIFQRLRYETITLQSDHCLDPGIPVNGQRH 638
 Db 827 GEYHGTQVPOPLISTSVYLLFTSTKSHSDIFQRLRYETITLQSDHCLDPGIPVNGQRH 886
 QY 639 GNDPFGALVTFSCDSGYTLDGSEPLECEPNFQMSRALPSCALCGPFIQSSGTTILSPG 698
 Db 887 GGFPGIISTVTFSCDPEYITLSDDEPLVCERNQWNAHLPCDLCGVIQKSGTIVLSPG 946
 QY 699 FPDYFNNLNCWTIETSHGKGVFTFHTHLSGHDYLLITENGSTFOPLRQLTGSRLP 758
 Db 947 FPDYFNNLNCWTIETSHGKGVFTFHTHLSGHDYLLITENGSTFOPLRQLTGSRLP 1006
 QY 759 APIISAGIYGNFTQVRIISPSMSYEGFNITTFSEYDLEPCEBPVAVYSIRKGLQFVGD 818
 Db 1007 HTIKAGIFGNFTQVRIISPSMSYEGFNITTFSEYDLEPCEBPVAVYSIRKGLQFVGD 1066
 QY 819 TLFFSCPGVRLBGTARTITLIGARRRMSPLPRCVACGNSVTGCTLLSPFNPVNV 878
 Db 1067 SLTFSCTGLRLEGATVLTLLGGRRWMSAPLRCAVACASVKNNGTLLSPFNPVNV 1126
 QY 879 NNHECIYSIQTPQKGIQLKARAFELSEGVLVKYDGNNSARLLGVFSHEMNGVTLNS 938
 Db 1127 NTHECIKYITBEAGKHILHTRSPQLPEGDTLKYVDKSDSSRPLGFTYXNELLGLTINS 1186
 QY 939 TSSSLMIDFTTDAENYSKGFELHPSFELKCEDPGFKPKGYKVDHSGAGSSVSSCD 998
 Db 1187 TSNHLMLEFNTNSGDDTQGFQLTYSFDLVKCDPGIPNNGYRIRBEGHFTDVTVLVSCN 1246
 QY 999 PGVSLRSGSEILCLSGERRTMDRPLPTCVACGCGTVGVEGVLSPGYAPAEHNLNCI 1058

Db 1247 PGYAMHSGNLTJCLSGDRRWMDKPLPSCIAECGQIHAATSGRIILSPGYAPADYNNHCT 1306
Qy 1059 WTEEAAGCTIGIHEHVPTTEEVHVDLRIWQGVBSGVLLKELSGPALPKDIHSTNSV 1118
Db 1307 WTEAHPGKTIISHFIVFTEMAHDLTKWMDGPVSDILLKEMSGSALPEDIHSTNSLT 1366
Qy 1119 LQFSTDFTSKQGFALQFVSSTATSCNDPGI PONSRSRSGDSWEADSTVFOCDPBYALOG 1118
Db 1367 LQFSDDFIISKSGFSIQFSTSIATCNDMPQNGTRGDSWEADSTVFOCDPBYALOG 1426
Qy 1179 SAEISGVKLENFPFQOPSPPTCIAQCGDLTGPSCVILSPNPEBPYKCEMDKVTVP 1238
Db 1427 QAKITCVQLMNRFQOPBPPTCIAACGDLTGPACVILSPNPPCYPKCEKCDMKVKNP 1486
Qy 1239 DYVALVFNIFNLIEGYDFLHIYDGRDSLPLISGFSQOLPGRTESSNSLFLAFPSDA 1238
Db 1487 DEVIALIFKSPFMESYDFLHIYEGEDSNPLISGFSQOLPGRTESSNSLFLAFPSDA 1546
Qy 1299 SVSNAGFYVDYENPESCDFBGSIKNGTRVSGDLKLGSSVYVYCHGGEVEGTSTSCI 1358
Db 1547 SVGLSGFALIEFKEKREACFPDGNIMNGTRVGTDPKLGSTITTCDSGKILDPSSITCV 1606
Qy 1359 LGPDKPVMNNRPVCTAPCGGOYVSGDVLSPPYQNTYSGQICLYFVTPKDYVVF 1418
Db 1607 IGADKPSMDVYLBSCNAPCGGOYVSGDVLSPPYQNTYSGQICLYFVTPKDYVVF 1666
Qy 1419 QPAFHIALNDVVEVHDSQHSRLSSLSGSHTEGSLPLATSNQVLIKFSAKLAPARG 1478
Db 1667 QFAYFQIALNDLAEFLPDGTHAQLRLSSLSGSHTEGSLPLATSNQVLIKFSAKLAPARG 1726
Qy 1479 FHFVYQAPRTSAQOCSSVPEPRYKRLGSPFSGAIVREPCNSGVALQSSPETECLPVP 1538
Db 1727 FHFVYQAPRTSDQOCSSVPEPRYKRLGSPFSGAIVREPCNSGVALQSSPETECLPVP 1786
Qy 1539 GALAQMNVASPTCVPCGGLTERGTILSPFEPFLNLSINCMWKIVVEGAGIQOVV 1598
Db 1787 NALAQMNDTIPSCVPCGGLTERGTILSPFEPFLNLSINCMWKIVVEGAGIQOVV 1846
Qy 1599 SFVTEQWMD-SLEFDPQADNTVMLGSPGTTVPALINSTNOILYHFYSDISVSAAGH 1657
Db 1847 QFCHGABELGJPFELHHDGDVYAPRLGSPGTTVPALINSTNOILYHFYSDISVSAAGH 1906
Qy 1658 LEYKTVGLSSCPPEAVNSGVKTGERLYLVNDVVSFOCEPGYALQGHANHSCHMGTRRAN 1717
Db 1907 LEYKTVGLAACQEBALNSIKIGDRYLVNDVVSFOCEPGYALQGHANHSCHMGTRRAN 1966
Qy 1718 YPPEPLCIAQCGGVTEENEGVILSPGEPGNYPSNMDCSMKIALPVGCAHIQFLNFTSTEN 1777
Db 1967 YPPEPLCIAQCGGVTEENEGVILSPGEPGNYPSNMDCSMKIALPVGCAHIQFLNFTSTEN 2026
Qy 1778 HDYIEINSGPYETSRMMGRFSGSELBSLSTSHETTVYFHSHSQNRPGFKLEYQAYEL 1837
Db 2027 HDYIEINSGPYETSRMMGRFSGSELBSLSTSHETTVYFHSHSQNRPGFKLEYQAYEL 2086
Qy 1838 QCEPDPPEPANGIVRGAGVVGQSVTPECLPGYOLGHPVLTCQGHNTNNMHPYKCEV 1897
Db 2087 QCEPDPPEPANGIVRGAGVVGQSVTPECLPGYOLGHPVLTCQGHNTNNMHPYKCEV 2146
Qy 1898 PCGNGTSSNGTYSGFPSPYSSODCWLITVPIGHGVRLNLSLQTEPSGDEITIMD 1957
Db 2147 PCGNGTSSNGTYSGFPSPYSSODCWLITVPIGHGVRLNLSLQTEPSGDEITIMD 2206
Qy 1958 GPOQIAPRLGVFTSRMAKKTVOSSNQVLLKFRBDAATGIFALIPASVAPLTKCPPTIL 2017
Db 2207 GPOQIAPRLGVFTSRMAKKTVOSSNQVLLKFRBDAATGIFALIPASVAPLTKCPPTIL 2266
Qy 2018 PNAEVTENEBENIGDIVRCLPGFTLVGNELTCKLGTYLQFEGPPICEVHCPTNEL 2077
Db 2267 PNAEVTENEBENIGDIVRCLPGFTLVGNELTCKLGTYLQFEGPPICEVHCPTNEL 2326
Qy 2078 LTFDSTGVILSOGYPSGYFQFCSMIVREBPYINSLYEVYFLSKQYDFEPIFDGSPGQ 2137
Db 2327 RKGSSGVILSPGYFNSQFCSMIVREBPYINSLYEVYFLSKQYDFEPIFDGSPGQ 2386

Qy 2138 SPLIKALSGVAPLIVTSSNSGVYLRMSDHAIVNKGFRIRYASAPYCLPAPALHGFIL 2197
Db 2387 SPLIVYLSGNHTQSNFTSRNSQILYRMSTIDHATSKKGRIRYAPAYCSLTHPLKNGGIL 2446
Qy 2198 GQTSSTOPGSGIHFECNAGYRLVGHSMALICTRHPOGYHLSSEAPILCOALSCGLPEAPKNG 2257
Db 2447 NRTAGAVGSKVHYFCKPGRMYGHSNATCRBNPLGMYQWDSILPLCQAVSCGIPESRNG 2506
Qy 2258 MVEGKEYTVGYAVVSCSEGYHLOGAELTACDLDTGLMSNNVNPQCYVTPCPDVSSIS 2317
Db 2507 SFTGNEFTLDSKVYVECHGEFLESQQLTAVCOEDGLMSNGKPMCKPVACPSIEQL 2566
Qy 2318 VEHGRNRLIFETQYQFQALMLICDPGYVYTGQRYARCQANGMSLGDSTPCTRILISGE 2377
Db 2567 SEHIVRLVSGSLNEGAQVLLSCPGYLLBEMRLRLQCANQNTWNIIGBERECRVISSCS 2626
Qy 2378 LPIPMGRHIGTLVYAGATAFSCNSGTYLVGSRVRECMANGMSGSEVRCIAGHCTPE 2437
Db 2627 LSPPMGNKIGTLVYAGATAFSCNSGTYLVGSRVRECMANGMSGSEVRCIAGHCTPE 2686
Qy 2438 PIVNGHINENYSGSVVYQCNAGFRLIGMSVRLICQODHMSGKTPCVPITTCGHPNP 2497
Db 2687 PIVNGHISGDSFSYRDTVYQCNPGFRLVGSVRLICLODHMSGKTPCVPITTCGHPNP 2746
Qy 2498 VNGLTQGNQFNLDVYKFCVCPGYABEGAAASQCLASQMSDMLPTCRILINCTDPGHQEN 2557
Db 2747 AHGFTNGSEFNLNDVNFCTCMTGYLLQGVSAQCRSNQMSPLPTCGVNVCSDPGFVEN 2806
Qy 2558 SVRQVHAAGPFRFSFGTVSVRCHNGFYLCTPVLSCQDCTMRPRRQCLLVSCGHPGS 2617
Db 2807 AIRHQGNPFSFEGMSILYHCKKGYLLSSALTCANLMPRSJPKCLATISCHPHGV 2866
Qy 2618 PPHSQMSDSTVGAIVRYSCIGRTLVGNSTRMCGLDGHTGSLPHCSGTSVGVCDPG 2677
Db 2867 PANAVLTGELFTYGAIVHYSCRGSESLIGNDTRVCOEDSHMSGALPHCTGNPPFCDDPG 2926
Qy 2678 IPAHGRIRLGDSPFDGTVWRSCFEGHYLRGSEERTCOANGMSGSOPECVYISCGNPSTP 2737
Db 2927 TPAGHSRGLDFTKSLRFSCEMGHQRLRSPERTCLLNSWSGLQVCEAVSGSGNGETP 2986
Qy 2738 SNARVPSDGLVFSSIVYECREGYATGLSRHCSVNGTWSGSDPECLVINCDDPGIPA 2797
Db 2987 TNGMIVSDGILFSSVYIYACWEGYKTSGLMTRCTANGTWGAIBPCTYIISGDBGTILA 3046
Qy 2798 NGLRLGNDFRNKTVYQCVPGYMMESHRYVSUSCTKDTWNGTKPVCKALMKCPPLIP 2857
Db 3047 NGIOPGTDFTFNKTVSYQCNPGYMEAVTSATIRCTDGRMNSKPVCKAVLCPQPPVQ 3106
Qy 2858 NGKVVGSDPFMNGSSVYACLEGYQLSLPAYFTCEGNSWTGELPOCPVPCGDPGVPSRG 2917
Db 3107 NGTVGSDPFMNGSSVYACLEGYQLSLPAYFTCEGNSWTGELPOCPVPCGDPGVPSRG 3166
Qy 2918 RREBRGFSYSSVFSCHPPLVLVGSPPRRCQSDGTWGSTOPSCI-DPTLTTCADPGVPQ 2976
Db 3167 RLSGKSFYKSEVFFQCKSPFIIYVGSRRVCAQADGTWMSGIOPCTICIDPANHCPDGTTH 3226
Qy 2977 FGIQNNQGVQVSTYLFRCQKGYLLQGSTTRCLDNLTWMSGTRPPCVCVPHKCRQETPTH 3036
Db 3227 FGIQNNQGVQVSTYLFRCQKGYLLQGSTTRCLDNLTWMSGTRPPCVCVPHKCRQETPTH 3286
Qy 3037 ANYGALDLPMSGYTLI 3052
Db 3287 ADVRAIDLPTFGYTLV 3302

RESULT 12
AAE20789 standard; protein; 3100 AA.
AAE20789;
AC
XX
XX
01-JUL-2002 (first entry)

QY 1382 YVSGDVVLSBNYPONTYSQOICLYPTVPKDYVVFQGFAPFHTALNDVVEVHDSQHS 1441
 DB 1165 YTGSEGVVLSBNYHANTAGQICLYSTVPEKVFVFGQFALFOTLNDLBLEFHTAQ 1224
 QY 1442 RLSSLSGSHSGESLPLATSNQVLIKPSAKGLABARGFHYQAVPTSTATQCSVDEPR 1501
 DB 1225 RLSSLSGSHSGEFTPLATSNQVLIKPSAKGLABARGFHYQAVPTSTATQCSVDEPR 1284
 QY 1502 YGKSLGSDGVGALVRECNCGVYALQSPLECLPVGALAAQNVSAFTCVVPCGNLT 1561
 DB 1285 YGRITGSEFSGSIVREKXNPGYLLQGSTALHCQSVNALAQNMDTIPSCVPCPSGNFTQ 1344
 QY 1562 RRGITLSPGPEPEPLNSLNCWKIVPEAGAGIQIOVSPFTEQWMDSELEFDGADNTVM 1621
 DB 1345 RRGITLSPGPEPEPEPLNSLNCWKIVPEAGAGIQIOVSPFTEQWMDSELEFDGADNTVM 1404
 QY 1622 LGSFSGTTVPALNLSNQLYHFYSDISVSAAGFHELYKTVGLSSCPREPAVPSNGK 1681
 DB 1405 LGSFSGTTVPALNLSNQLYHFYSDISVSAAGFHELYKTVGLSSCPREPAVPSNGK 1464
 QY 1682 ERYLVNDVSPGCEGVALQGHAIQCMPTVRMNPPLCTAACGGTVEHEGVILSP 1741
 DB 1465 ERYLVNDVSPGCEGVALQGHAIQCMPTVRMNPPLCTAACGGTVEHEGVILSP 1524
 QY 1742 GPEGNYSNDQSKIALPVGFGAHIOPLNFSTEPNHDYIEIRNGPYETSRMMGRPSGE 1801
 DB 1525 GPEGNYSNDQSKIALPVGFGAHIOPLNFSTEPNHDYIEIRNGPYETSRMMGRPSGE 1584
 QY 1802 LBSLSLSTHETTVYFHSDSQNRPGKLEYQAYEIOECDPPEPPANGIVRGAYVGS 1861
 DB 1585 LBSLSLSTHETTVYFHSDSQNRPGKLEYQAYEIOECDPPEPPANGIVRGAYVGS 1644
 QY 1862 VTECELPYQOLTGHPVLTQCHGTRNMDHPLPKCEVPCGGNITSSNGTVSPPEFPYS 1921
 DB 1645 VTECELPYQOLTGHPVLTQCHGTRNMDHPLPKCEVPCGGNITSSNGTVSPPEFPYS 1704
 QY 1922 SQDCWMLITVPIGHVRLNLSLQTEPSGDFITIMDPOQTARLGVFTSMMAKTVOGS 1981
 DB 1705 SQDCWMLITVPIGHVRLNLSLQTEPSGDFITIMDPOQTARLGVFTSMMAKTVOGS 1764
 QY 1982 SNOVLKFRHDAATGCFALAFSAVPLTKCPPTLIPMAEVTENEENIGDITVRCLP 2041
 DB 1765 SNOVLKFRHDAATGCFALAFSAVPLTKCPPTLIPMAEVTENEENIGDITVRCLP 1824
 QY 2042 GFTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 2101
 DB 1825 GFTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 1884
 QY 2102 WLVREPDYINISLTVYFLSEKQYDEFEIJDGSGSPILKALSGNYSAPLIVTSSNSV 2161
 DB 1885 WLVREPDYINISLTVYFLSEKQYDEFEIJDGSGSPILKALSGNYSAPLIVTSSNSV 1944
 QY 2162 YLRMSDHLVNRKGFIRYASAPYCSLPARLHFIIGOSTOPGSHHGCAVRLVGH 2221
 DB 1945 YLRMSDHLVNRKGFIRYASAPYCSLPARLHFIIGOSTOPGSHHGCAVRLVGH 2004
 QY 2222 SNAICTRHPQGYHLMSEALPLCOALSCGPEAPKNGMFGKETYVGTAKVSCSEGYH 2281
 DB 2005 SNAICTRHPQGYHLMSEALPLCOALSCGPEAPKNGMFGKETYVGTAKVSCSEGYH 2064
 QY 2282 AGAEATAECLDTGLMGNRVNPPQCVPTCPDVSSISVEHGRWLLIFETOYQFOALMLT 2341
 DB 2065 AGAEATAECLDTGLMGNRVNPPQCVPTCPDVSSISVEHGRWLLIFETOYQFOALMLT 2124
 QY 2342 DPGYVYTGQVIRCOANGKMSLGDSTPTCRITISCEGLPIPRGHRIGTISVGAARLSC 2401
 DB 2125 DPGYVYTGQVIRCOANGKMSLGDSTPTCRITISCEGLPIPRGHRIGTISVGAARLSC 2184
 QY 2402 NSGYTLVGSRVCECMANGLMSGEVYCLAGHGTPEPIYNGINGENYSGSVYQCN 2461
 DB 2185 NSGYTLVGSRVCECMANGLMSGEVYCLAGHGTPEPIYNGINGENYSGSVYQCN 2244

QY 2462 GFTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 2521
 DB 2245 GFTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 2304
 QY 2522 MEGARSQCLASQCSMDLPTCRITINCTDPGHQENSVYQVASGPHRSFGTTSYRCN 2581
 DB 2305 MEGARSQCLASQCSMDLPTCRITINCTDPGHQENSVYQVASGPHRSFGTTSYRCN 2364
 QY 2582 HGFYLLGTPLVLSQGDGTDRPPOCLLVSCGHPSPPHSOWSGSYTVGAVVYSCIK 2641
 DB 2365 HGFYLLGTPLVLSQGDGTDRPPOCLLVSCGHPSPPHSOWSGSYTVGAVVYSCIK 2424
 QY 2642 RTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 2701
 DB 2425 RTLVGNEILTCAGTYLQEPGPPICEVHCPFTNELTSTGTYLSQSGSYPOFQCS 2484
 QY 2702 GHVLRSSERTQANGSGSGSPGEGVILSCGNPTSPNARVYFSDGLVSSSIYECREG 2761
 DB 2485 GHVLRSSERTQANGSGSGSPGEGVILSCGNPTSPNARVYFSDGLVSSSIYECREG 2544
 QY 2762 YATGLLSRHGCVNGTWSDECLVINGDPIRANGRLGANDFRYKTVTYQCPGYM 2821
 DB 2545 YATGLLSRHGCVNGTWSDECLVINGDPIRANGRLGANDFRYKTVTYQCPGYM 2604
 QY 2822 MESHVSVLSCDKRTMNGTKRVCKALCKPPELLPNKVVYSDPMGSSVYACLEGY 2881
 DB 2605 MESHVSVLSCDKRTMNGTKRVCKALCKPPELLPNKVVYSDPMGSSVYACLEGY 2664
 QY 2882 LSLPAVFTCEGNGSTGELPOCFPVPCDPOGVPSRGRREDRFSYSSVSFCHPLVLY 2941
 DB 2665 LSLPAVFTCEGNGSTGELPOCFPVPCDPOGVPSRGRREDRFSYSSVSFCHPLVLY 2724
 QY 2942 GSPRRFCSQDGTWSGTOSCLDPLITTCADGCVPOFGIQQNSQYQVSTVLFRCQKYL 3001
 DB 2725 GSPRRFCSQDGTWSGTOSCLDPLITTCADGCVPOFGIQQNSQYQVSTVLFRCQKYL 2784
 QY 3002 LQGSTTRICLANLWTSGLQTECIRHACQOPETPAHADYALDLPFGYTLV 3052
 DB 2785 LQGSTTRICLANLWTSGLQTECIRHACQOPETPAHADYALDLPFGYTLV 2835

RESULT 13
 ID AAE20901 standard; protein; 3100 AA.
 AC AAE20901;
 DT 01-JUN-2002 (first entry)
 DE Human C3b/C4b complement receptor like protein #2, alternative version.
 KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
 KW transplant rejection; autoimmune disease; ischemic condition; necrotic;
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
 KW infertility; vasodilator; obesity; cardiac.
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 418
 FT /label= unknown
 FT Misc-difference 492
 FT /label= unknown
 FT Misc-difference 617
 FT /label= unknown
 FT Misc-difference 726
 FT /label= unknown
 FT /note= "Encoded by CTR"
 PN W0200210199-A2.

XX 07-FEB-2002.
 PD 24-JUL-2001; 2001WO-US023232.
 XX 02-AUG-2000; 2000US-0222504P.
 XX 28-NOV-2000; 2000US-00728787.
 XX (AMGE-) AMGEN INC.
 PI Welcher AA, Elliott GS;
 DR WPI; 2002-303934/34.
 DR N-PSDB; AAD33320.
 XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
 PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
 PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
 PT sclerosis.
 PS Claim 13; Page 239-251; 251pp; English.
 XX The invention relates to a nucleic acid encoding a novel C3b/C4b
 CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
 CC polypeptide and nucleic acid molecules may be used to treat, prevent,
 CC ameliorate, diagnose and/or detect diseases such as immune system
 CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
 CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's
 CC disease), ischaemic conditions, metabolic disorders (e.g. obesity and
 CC diabetes) and infertility. The invention is useful in gene therapy. The
 CC present sequence is human C3b/C4b complement receptor like protein.
 CC alternative version. Note: The present sequence is stated to be the same
 CC as that referred to as SEQ ID NO:7 (AAE20789) shown in figure 2 of the
 CC specification. However the sequences differ at position 726
 XX Sequence 3100 AA:
 SQ
 Query Match 63.3%; Score 10755; DB 5; Length 3100;
 Best Local Similarity 64.5%; Pred. No. 0;
 Matches 1853; Conservative 429; Mismatches 545; Indels 44; Gaps 5;
 190 TLTVGDDGGGDDGQDQTVLYMSQNAACSDSBRHTGSRIPR---SMSDIDRQKTVLVEICRDI 246
 1 TLTVGDDGAGKVGDTSSVLYV-----LTGSSVPLDLYMSNQMLH-----L 40
 247 SSDDARSGVSKSPKTSNAVELVAPG-----TEIEGSCGDPGIPAYGRGSRFHHGDT 301
 41 QSDDS-----IGSPGRKAVYQELIEKGGCDPGIPATGKRGSSFLHGD 84
 302 LKFECPAPAFELVGGKAITCQKNNQMSAKKPCVSCFPNFTSPGVVLSPPVPEYGNHL 361
 85 LTFECPAPAFELVGGKAITCQKNNQMSAKKPCVSCFPNFTSPGVVLSPPVPEYGNHL 144
 362 HCVWLLIARPSRRIHLAFNDIDVPPQPLVYKGGATRAEAVLGTTFEGNQLPSSITSSGH 421
 145 NCWVLLIASEPGRRIHLAFNDIDVPPQPLVYKGGATRAEAVLGTTFEGNQLPSSITSSGH 204
 422 VARLEFQDHSHTGKRGFNITFTFRHNECPDPGVVNGKRGFGLSLQSSISFLCDSGFL 481
 205 IVRLFEQDHSHTGKRGFNITFTFRHNECPDPGVVNGKRGFGLSLQSSISFLCDSGFL 264
 482 GTGSEFTTVCVLEKGSVVMNSAVLACEAPCGHLLTSPGPGFYKDALSCAMVI 541
 265 KTQSESEFTTVCVLEKGSVVMNSAVLACEAPCGHLLTSPGPGFYKDALSCAMVI 324
 542 EAQCGYPIKTFEDPKFEVNVDTLEVRDGRYSAPLGVYHGTQVPPFLITSXVYLFL 601
 325 EAKGHSKIKTFEDPKFEVNVDTLEVRDGRYSAPLGVYHGTQVPPFLITSXVYLFL 384
 602 STDGSHSDIGFQLRYETITLQSDHCLDPGIPVNGQGRHNDPFYVALVTFSCDSGYTLSDG 661

DB 385 TTDSNRSSIGFLIHYESVTLESDCLDPGIPVXKHRRGGDFGIRSVTFSCDPGYTLSD 444
 662 EPLECEPNQMSALPSCEALCGGFIQSGSGLTSLSPGPPDFYNNINCTMIIFESHGKV 721
 445 EPLVCERNQMNALPSCDALCGGYIQGKGYLSPGPPDFYNSLXKWTIVESHGKV 504
 722 FFFHTFHLSEGDYLLITENGSTFOTPLROLTGSRLLPAPISAGLYGNTAQRVFIQFSM 781
 505 QMIFHTFHLSESHDYLITEDGSPSEFVALVTSVLPHTIKALFGFNTQALFISDFSI 564
 782 SYEGFNITFSEYDLPECEBEVPAYSIKGLQGVGDTLTFSCPPGYRLGARITCLG 841
 565 SYEGFNITFSEYDLPECDPDGVPAPFASRIIFHGVGDSLTFSCFLGYRLGAXKLTCLG 624
 842 RRLMSPLPRCAVECGNSTVTGQTLSSNPNVNNHNECTYSIOTQKGIOLKARA 901
 625 GRWVMSPLPRCAVECGNSTVTGQTLSSNPNVNNHNECTYSIOTQKGIOLKARA 684
 902 FELSEGVLYKYDGNNSARLLGVFSHEMVGTLNLSSTSLWLDFTDAENTSKEFELH 961
 685 FOLFEBDTLKYVQKXSSRPLGTFTKNELGLILNLSHMXLLEFNTGSDTDGQFLT 744
 962 FSSPELICKDPEPTPKRYVHDEHPAGSSVSFSCDPGYLRSSEELLCLSGERRTWR 1021
 745 YTSFDLYKCBDDPGIPVNGYRIIRDEGHFTDVLVYSCNPGYAMGSLTCLSGDRRWMDK 804
 1022 PLPTCAVECCGTGREGVSGOVLSPGYAPRBEHNLNCWTLEAAGCTIGHPLVPDEEV 1081
 805 PLPSCIAECGQIHAATSGILSPGYAPRBEHNLNCWTLEAAGCTIGHPLVPDEEV 864
 1082 HDVLRIMDGVESGVLLLEKESGALPKDILSTNSVLYQSTPFTSKQFALQFSVSTA 1141
 865 HDLKWKGVDVSDILKESGALPKDILSTNSVLYQSTPFTSKQFALQFSVSTA 924
 1142 TSCNDPGIPONGSRSGDSWAGSTVQCDPGYALQGSABISCVKLENRFWQSPPTCI 1201
 925 ATCNDPMPONGRYSRSGADTVTFQCDPGYALQGSABISCVKLENRFWQSPPTCI 984
 1202 APGCGDLTSGSVILSNTPPEPYPGECMCKMTVSPDYIALAVENFLEPGDPLH 1261
 985 AACGGMITGAGVILSNTPPEPYPGECMCKMTVSPDYIALAVENFLEPGDPLH 1044
 1262 DGRDLSLPLGSGFSGOLPGRISSSNSLFLAFRSDASVNAAGFVIDYTNPRESCDPG 1321
 1045 EGEDSNPLGSGFSGOLPGRISSSNSLFLAFRSDASVNAAGFVIDYTNPRESCDPG 1104
 1322 SIKNGTRVSGDLKSGSVITYYCHGYVEGISTLCLGADGKRVNNPAPVCTAPCGGQ 1381
 1105 NIMNGTRVSGDLKSGSVITYYCHGYVEGISTLCLGADGKRVNNPAPVCTAPCGGQ 1164
 1382 YVSGDGVVLSPPNTPONTSQICLYPYTVYKDYVVFQGFAPFHTALNDVVEVDHSHQS 1441
 1165 YTSSEGVVLSPPNTPONTSQICLYPYTVYKDYVVFQGFAPFHTALNDVVEVDHSHQS 1224
 1442 RLTLSSLSGSHGSLPLATSNQVLYIKPSAGLAPARGFHFVYQAVPTSAQSSVPEPR 1501
 1225 RLTLSSLSGSHGSLPLATSNQVLYIKPSAGLAPARGFHFVYQAVPTSAQSSVPEPR 1284
 1502 YGRKLSGDFSGVAILVREPCNSGYALQSPRIECLPYVGALQMNVAAPTCVPCGNL 1561
 1285 YGRKLSGDFSGVAILVREPCNSGYALQSPRIECLPYVGALQMNVAAPTCVPCGNL 1344
 1562 RRGITLSPGYPPEYGNLNCIMKIIVYEGGIDQIVISFATEQWDSLETHDGDVAPR 1621
 1345 RRGITLSPGYPPEYGNLNCIMKIIVYEGGIDQIVISFATEQWDSLETHDGDVAPR 1404
 1622 LGSFGTTPALNLSNOLYLHFSDISVSAGFHELYTVGSLSCPEBAVNSNGVKTG 1681
 1405 LGSFGTTPALNLSNOLYLHFSDISVSAGFHELYTVGSLSCPEBAVNSNGVKTG 1464
 1682 ERYLVNDVVSFOCEPGYALQGHAIISGMPGTVRMNPBPPLCIAQCGTYVEGEGVTLSP 1741
 1465 DRYVNVNDVVSFOCEPGYALQGHAIISGMPGTVRMNPBPPLCIAQCGTYVEGEGVTLSP 1524

QY 1742 GEPGVPSNMDCSKIALPVGFAHIOFLNSTEPNDYITRNQPYETSRMGRFEGSE 1801
 Db 1555 GPPGSGYNNLDCITWIKISLPIGYGAHIOFLNSTEPANDFLEIQNPHTSMICQFSGTD 1584
 QY 1802 LPSLLSTSHETTYVFHSDHSONRPGFLEAYQAYELOECPDPPEFPANGIVRAGAYNGOS 1661
 Db 1585 LPAALLSTHETHTLHFHSDHSONQNGFLAQAELQNCBPFPFGNGYMINSTYSGOS 1644
 QY 1662 VTEFCLRGYQLTGHVPLTQCHGTRNNDHP.LKCEVPCCGNITSSNGTIVYSPGFPDYS 1921
 Db 1445 VSFECYPRYLLIGHPVLTCHGIRNNMNYPPPRCDACGVYVTSQNGTIVYSPGFPDYP 1704
 QY 1922 SODCVMLITVPIGHVRLNLSLQTEPBGDEITLWDEPQOAPRLGVFTSMKATYQSS 1981
 Db 1705 LKDCIMLITVPBGGVYINFTLQTEAVNDIYAVMDPDQNSPOLGFSNGNTLETAYSS 1764
 QY 1982 SNOVLLKFRHDAATGGIFAIAFSAVPLTKCPPTILPNAEVTENEENFNGIDIVYRCLP 2041
 Db 1765 TNQVLLKFRHSDFSNGFEVLFNFHAFQKKQCPRAVQAEMLTEDDDFEIGDVPYKQCHP 1824
 QY 2042 GFTLVGNEILLTKCTGTLOFEGEPPIEGEVHCFPTNELLTSTGYLLSGSYGSPQRTCS 2101
 Db 1825 GYTLVGTDLTKCLSSQLQFEGSLPTCEAQCANEVRTSGSVILSPGYGNVFNSTQCS 1884
 QY 2102 WLVRBPDYINISLVEYFLESEKQDEFEIFDGPBGSGPLLKALSGVNSAPLIYTSNSV 2161
 Db 1885 MSKIVENNYNITIFVDIFQSEKQDFAEIVDGGSGGSPLIYVLSGNTGDSNTTSRNL 1944
 QY 2162 YLRWSSHAMNKRKGFRIYAPYCSLPRAPLHGFILQSTQSTQPGSIHFCNCAGYRLVGH 2221
 Db 1945 YLRSTHATSKKGFRIYAPYCSLPRAPLHGFILQSTQSTQPGSIHFCNCAGYRLVGH 2004
 QY 2222 SMAICTHPPGYHLMSEAIPLCOALSCGLPEARKNVFEKYEYTGKAVYSGSEGHQ 2281
 Db 2005 SNAICTRNPGLGMOWBSLTP.LCOAVSCGIPESPNSSFTNEFTLDSKVVEYEHGEPKLE 2064
 QY 2282 AGABATACLDLDTGIMSNRNPPOCVPTCPDVSSISVEHGRMLIFETQYFOQAQMLIC 2341
 Db 2065 SSOQATFVCOEDGIMSNKGRPTCKPVACPSIEAQSEHNTWRLVGSINTEYAOVLBC 2124
 QY 2342 DRYVYVTCQVIRCOANGKMSLDSTPTCTGIISSCGELPIPNGRHRTGTSVYGAINIIFSC 2401
 Db 2125 SPGYVLEGWMLRFOANGTWNIDBERSCRVISCGISPPPNKIGTLTVGATATIFTC 2184
 QY 2402 NSGYTVGSRVRECANGMLWSGSEVRCLAGHCTPPIVNGHNGENYSGRVVYQCA 2461
 Db 2185 NTGYTVGSHVRECLANGMLWSSETRCLAGHCSPPPIVNGHISGDFSTRTIVYQCNP 2244
 QY 2462 GFRLLGMSVRIQQDDHMSGKTPFCVPTTCGHPNGFVNGLTQGNQPNLNDVVKFVCPNGY 2521
 Db 2245 GFRLLVGTSVRICTQDHRKMSQTPVCPITCGHPNGHAFGTNGSEFNLDVNFCTNGY 2304
 QY 2522 MEGARSOCLAGSGOMDLPTCRILINCDBPGHGENSVQVHAASGHRSEFTGYTSYRN 2581
 Db 2305 LLAGVRAQCRSGNQSSPLPTCRVNCSDPGFVENAIGHGOONFESFEYSILYHXK 2364
 QY 2582 HGFYLLGTPLYSCQGDGTWDRPRPQCLLYSCGHPGSPPHSGMSGDSYTVGAVVRSCTGK 2641
 Db 2365 KGFYLLGSSALTCMANGLWDRSLPKCLASCGHPVPAVAVLTLGELFTYGAVVHSCRS 2424
 QY 2642 RTLVGNSITMGLDGHWTGSLPHSCSTSVGCGDPEI.PAHGIRLGDSPDGTVMRPSCEA 2701
 Db 2425 ESTLIGDITVCOEDSHMSGALPHCTGNNPFGCDPPTPAHGRSLGDFXTKSLTRFSCEM 2484
 QY 2702 GHVLRSSERTCOANGMSGSGOPEGVISCANPGTSPSNARVVFSSGIVSSIVVECREG 2761
 Db 2485 GHQLRSSPERTCLANGMSGLOPVCEAVSCGNPGITPNTMITVSSGILFSSSVIYACWEG 2544
 QY 2762 YVATGLSHHCGVNGTWTGSDPECLVINGCDPQIPANGLRLNDRFRYNTVYQCVPGYM 2821
 Db 2545 YKTSIGIMTHCTANGTWTGTADPCTIISCGDPGLTANGI.QFCTDTPFKNTVSYQCNPGYV 2604

QY 2822 MESHRSVLSCTYDRTWNGTKPVCKALMCKPPPLIPNGKRVGSDFMWSSVTYACLEGYQ 2881
 Db 2605 MEAVTSATIRCTKDGHRMNSKPKCKAVLCPQPPPVONGTVESGDFRMWSSISVSCMDGYQ 2664
 QY 2882 LSLPAVFTCEGNSWNGELPOCEPVPCGDPVPSKGRREDRGRFSRSSISFSCHPPLUV 2941
 Db 2665 LSHSALISCEGKWMGELPQCLPVFCGDPGIPABGRISGKSFYKSEVFFQCKSPFLIV 2724
 QY 2942 GSPRRPCQSDGTWSTGTPSCIDPTLTTCADPCVPQFQIONNSQGYOVGSTVLFRCKGYL 3001
 Db 2725 GSRBRVQADGTWMSGI.QPTCIDPANTCPDPGTPHFGIONSRSRGYEVGSTVFRCKGYH 2784
 QY 3002 LOGSTRTCLPULTMSGTTPDVCVPHHCROPEFTPTANALDLPMSNGYTLI 3052
 Db 2785 IGGSTTRCLANLWMSGIQTCECIPACRQPEFPAHADVAIDLPTFGYTLV 2835

RESULT 14
 AAE20788
 ID AAE20788 standard; protein; 3095 AA.
 XX
 AC AAE20788;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Rat C3b/C4b complement receptor like protein.
 XX
 KW Rat; C3b/C4b complement receptor-like molecule; immune system disorder;
 KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
 KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;
 KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
 KW transplant rejection; autoimmune disease; ischemic condition; nocturnal;
 KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
 KW infertility; vasodilator; obesity; cardiant.
 XX
 OS Rattus rattus.
 XX
 FH Key
 FT Misc-difference 196
 FT /label= Unknown
 FT /note= "Xaa can be any amino acid"
 FT Misc-difference 847
 FT /label= Unknown
 FT /note= "Xaa can be any amino acid"
 PN W0200210199-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 24-JUL-2001; 2001WO-US023232.
 XX
 PR 02-AUG-2000; 2000US-0222504P.
 PR 28-NOV-2000; 2000US-00728787.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Welcher AA, Elliott GS;
 DR WPI, 2002-303934/34.
 DR N-PSDB; AAD33319.
 XX
 PT Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
 PT acid molecule, useful for treating, preventing and diagnosing rheumatoid
 PT arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
 PT sclerosis.
 XX
 PS Claim 13; Fig 3; 251pp; English.
 XX
 CC The invention relates to a nucleic acid encoding a novel C3b/C4b
 CC complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
 CC polypeptide and nucleic acid molecules may be used to treat, prevent,
 CC ameliorate, diagnose and/or detect diseases such as immune system
 CC disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory

CC arthritis, osteoarthritis, inflammatory joint disease, autoimmune
 CC disease, multiple sclerosis, lupus, inflammatory bowel disease,
 CC transplant rejection, nervous system disorders (e.g. Alzheimer's
 CC disease), ischemic conditions, metabolic disorders (e.g. obesity and
 CC diabetes) and infertility. The invention is useful in gene therapy. The
 CC present sequence is rat C3b/C4b complement receptor like protein
 XX

Sequence 3095 AA:

Query Match 63.1%; Score 10720; DB 5; Length 3095;
 Best Local Similarity 64.3%; Pred. No. 0;
 Matches 1843; Conservative 436; Mismatches 543; Indels 44; Gaps 5;

QY 195 DGGGDDGOKTVLWMSQNAQSDSPHTPGRIPE-----SSMGDIWRKWTYLETICRISSDA 251
 DB 1 DAGVGGDTRSLVLY-----LTGSSVPDLIVSMNQWLH-----LQSDDS 40
 QY 252 RSGSVRSKPTKSNVAVAPG-----TEIQSGCGDPGIPAYGRREGSRFHGDTLKEFC 306
 DB 41 -----IGSPGFAVYQIEIKGGCGDPGIPAYGRRTGSSFLHGDTLKEFC 84
 QY 307 QPAELVGGKAITQKNNQSAKKRPGCVSCFENFTSBSGVLSPTNEDYGNHLACVWL 366
 DB 85 QPAELVGERVITCQRNNQWGNKRPSCVSCFENFTASSGIIISPNTPEEGNMNMCVWL 144
 QY 367 ILAPESRIHLAFNDIVPEQDFDLVTKDGTAAEAPVLTGTSNQDLSSTSGHVALE 426
 DB 145 IISPSGRIRHLFNDPVEQDFDLAVDGDGIDITVLTGTSNVEVPPQLXSHIVALE 204
 QY 427 FQTHSTGKGFNTFTFRANECPPDPVPGNGKRGFSDLSQSSISFLCDEGLTQGS 486
 DB 205 FQSHSTGKGFNTFTFRANECPPDPVPGNGKRGFSDLSQSSISFLCDEGLTQGS 264
 QY 487 ETTTCVLKESGVNNSAVLRCEAPCGHLTSPSGTILSPGMPGYKDALSCAWYIEAOPG 546
 DB 265 ESTICIIQDGNVWVSTVPCREAPCGHLTASSGVILPPGMPGYKDALSCAWYIEAOPG 324
 QY 547 YPITITDRFTEVNYDTLEVRDRTYSAPLIGYHGVQVQPLISNTVYLTFSDKS 606
 DB 325 HSITITDRFTEVNYDTLEVRDRTYSAPLIGYHGVQVQPLISNTVYLTFSDKS 384
 QY 607 HSDIGPOLRYETITLQSDHCLDPCI PVNGQRHNDPVYALVTFSCSGYTLSDGEPLC 666
 DB 385 RASVGFILHESVTLSDSCLDPCI PVNGQRHNSNFGRTSTVSCDPGTYLSDDDLIC 444
 QY 667 EPNFQSRALPSCALCGGFIQSSSGTILSPGPDYPPNNLNTWIIETSHGKGVPTFH 726
 DB 445 EKNQNMHALPSCALCGGFIHGSQGVTLSPGPDYPPNNLNTWIIETSHGKGVQNMFH 504
 QY 727 TFFHESGHDYLLTENGSFQPLQLQGLSRLPAPISAGLYNFTNQRFTISDEMSYEGF 786
 DB 505 TFFHESGHDYLLTENGSFSEPPVARLTGSLVPHITKGLFENFTNQRFTISDEMSYEGF 564
 QY 787 NITSEYDLBCEPEEVPAYGIRKQLOGVGDTLTFSCFPGYREGTARITCLGGRRLW 846
 DB 565 NITAEVDLBCEPEEVPAYGIRKQLOGVGDTLTFSCFPGYREGTARITCLGGRRLW 624
 QY 847 SSPLRCAVACGNSVTGTGTLSPNFPVNNNNHCEYISIQTPGKGIQIKARAFELSE 906
 DB 625 SAPLRCAVACGNSVTGTGTLSPNFPVNNNNHCEYISIQTPGKGIQIKARAFELSE 684
 QY 907 GDVLYKVDGNNNSARLLGVFSHBMGVTLNSTSSLMDDPITTAENTSKGFLHPSSE 966
 DB 685 GDVLYKVDGNNNSARLLGVFSHBMGVTLNSTSSLMDDPITTAENTSKGFLHPSSE 744
 QY 967 LIKCEDETPKFGKGVHDEGHFAGSSVSPCDPGYSIRGSEBELICTLGERRTWDRPLPTC 1026
 DB 745 LYKCEDETPKFGKGVHDEGHFAGSSVSPCDPGYSIRGSEBELICTLGERRTWDRPLPTC 804
 QY 1027 VAECGGTVRGEVSGQVLSPGYADYEHNLNLCIWTIEABAGCTTGLAFLVPTDEEVDVLR 1086
 DB 805 VAECGGTVRGEVSGQVLSPGYADYEHNLNLCIWTIEABAGCTTGLAFLVPTDEEVDVLR 864

QY 1087 IMPGVESGVTLKELSGPALPKDLSTFNSVNLQSTDFPFTSKOGFALIOFSVSTATSCND 1146
 DB 865 VMDGPVDSNILLKEMSSALPEDIHSTNSLTQLQFSDFTISGSGEIOQSTSLASTSCND 924
 QY 1147 PGIPQNSRSGDSWEADSTVFOCCDPGALQGSAAEISCVKIEENRFWQSPSPPTCIACCG 1206
 DB 925 PGMPQNGRTYGDSEHPEDTITFOCCDPGALQGSAAEISCVKIEENRFWQSPSPPTCIACCG 984
 QY 1207 DLTPGSGVILSPNPPYPPGKCDMKVTVSPDVIALVAVNINLEBGFPLHIYGRDS 1266
 DB 985 NLTPGPAVILSPNPPYPPGKCDMKVTVSPDVIALVAVNINLEBGFPLHIYGRDS 1044
 QY 1267 LSPILGFSYSGOLPGRLESSNSLFLAFRSDASVNAAGFIADTENRREGCPDGSTKNG 1326
 DB 1045 NSPLIGFSYSGOLPGRLESSNSLFLAFRSDASVNAAGFIADTENRREGCPDGSTKNG 1104
 QY 1327 TRVGSDLKLGSSVTVYCHGGEVEGTSTLSCILGPDGRPMWNNRPVCTAPCGQGVYSD 1386
 DB 1105 TRIGTDFKLGSTVTVYCHGGEVEGTSTLSCILGPDGRPMWNNRPVCTAPCGQGVYSD 1164
 QY 1387 GVTLSPNPPYPPGKCDMKVTVSPDVIALVAVNINLEBGFPLHIYGRDS 1446
 DB 1165 GVTLSPNPPYPPGKCDMKVTVSPDVIALVAVNINLEBGFPLHIYGRDS 1224
 QY 1447 LSGHTEGSLPLATSNQVILKFSAGLAPARGFHFVYQAVAPRTSATQSSVPEPRYGRRL 1506
 DB 1225 LSGHTEGSLPLATSNQVILKFSAGLAPARGFHFVYQAVAPRTSATQSSVPEPRYGRRL 1284
 QY 1507 GSDPSVGAIVRFENSGSYALQGSPEIECLPVPGLAQMNVSAPTCVPCGSGNLTERRGTI 1566
 DB 1285 GSDPSVGAIVRFENSGSYALQGSPEIECLPVPGLAQMNVSAPTCVPCGSGNLTERRGTI 1344
 QY 1567 LSPGPEPYNLSLNCWKIVYPPBAGIQLQVYVSEVTEQWMDSELPFDGADNTVMGCSFS 1626
 DB 1345 LSPGPEPYNLSLNCWKIVYPPBAGIQLQVYVSEVTEQWMDSELPFDGADNTVMGCSFS 1404
 QY 1627 GTTVPALNSTNQLVHFYSDISVSAAGHLEKTKVGLSGCEPAVSPGVTGEREYL 1686
 DB 1405 GTTVPALNSTNQLVHFYSDISVSAAGHLEKTKVGLSGCEPAVSPGVTGEREYL 1464
 QY 1687 NDVVSFQCEBRYALQGHAMHISCPGTVRRNVPPLICIAOCGTVBEMEGVILSPGPN 1746
 DB 1465 NDVVSFQCEBRYALQGHAMHISCPGTVRRNVPPLICIAOCGTVBEMEGVILSPGPN 1524
 QY 1747 YPSNMDCKIALPVGEGAHITQFLNSTENNDYIETRNPYETSRMGFSSSELPLSL 1806
 DB 1525 YPSNMDCKIALPVGEGAHITQFLNSTENNDYIETRNPYETSRMGFSSSELPLSL 1584
 QY 1807 LSTSHETTVFHSQNSQNRPGFLEYQVAYELOECRDEPFRANGIVGAGVNGQSVTFEC 1866
 DB 1585 LSTSHETTVFHSQNSQNRPGFLEYQVAYELOECRDEPFRANGIVGAGVNGQSVTFEC 1644
 QY 1867 LPGYOLTGHPVLTQGHNTNRNDHPLPKCVPVPCGNTSSNGTVSPGSPVSSQDVC 1926
 DB 1645 LPGYOLTGHPVLTQGHNTNRNDHPLPKCVPVPCGNTSSNGTVSPGSPVSSQDVC 1704
 QY 1927 WLITVPVIGGVRLNLSLQTPESGDFITVWGQOQTAAPRLGVFTSRMAKTTVOSSNQVL 1986
 DB 1705 WLITVPVIGGVRLNLSLQTPESGDFITVWGQOQTAAPRLGVFTSRMAKTTVOSSNQVL 1764
 QY 1987 LKFRHDAATGIGFALPFAVAPLTKCPPTLLPAAEVTEBENIGDIYVRRCIPGTV 2046
 DB 1765 LKFRHDAATGIGFALPFAVAPLTKCPPTLLPAAEVTEBENIGDIYVRRCIPGTV 1824
 QY 2047 GNEILCTKLTGYLOFEGEPPPICEVHGTNELLDSTGVILISQSPGSPVPOFOTCSMLVRY 2106
 DB 1825 GNEILCTKLTGYLOFEGEPPPICEVHGTNELLDSTGVILISQSPGSPVPOFOTCSMLVRY 1884
 QY 2107 EPDYNTISLTYEYLRSEKQVDBEPIFDGSGQSPULLKALSGNYAPALITVSSNSVYLRWS 2166
 DB 1885 EPDYNTISLTYEYLRSEKQVDBEPIFDGSGQSPULLKALSGNYAPALITVSSNSVYLRWS 1944
 QY 2167 SDHAYNRKPKFIYSAFYCSLPAPRLHGFILGQTSQPGSSIHFGCAAGVRLVGHSMALIC 2226

```

Db      1945 TDHATSKGFKIRVAPYCSLSLTAKNGVLTAKGALGSKVQVFCRKGIMHSHNATC 2004
Qy      2227 TRHPGCHLMBEALFLCOALSCGLEPARKGNVFEKEYTGTAKVYSCSEGYHLQAGAEA 2286
Db      2005 RRPVGVYQWMSMAPLCOAVSCGIPLEAPGNCSFTGNFTLDSKYTYECNEGFKLDASQA 2064
Qy      2287 TAECIDTGLMSNRNVPQCVPTCPDVSSISVEGRMLIETQYQOQOQMLICDQGY 2346
Db      2065 TAVCOEDSLMNRKRPCTCKVPDPSIEGQISEHMLMVGSLNEGAQVLLSCSPGYF 2124
Qy      2347 YTGQRVIRCOANGKMSLSDSTPTCRILISCGELPIPNHGRIGILSVYATATFSCNGYT 2406
Db      2125 LOGQRLLQCANGTWNTEDRPRCKVISCGLSPFPNGKIGTLTMGATATFTCNTGYT 2184
Qy      2407 LVGRVRECMANGLMSSGEVACLAGHGTPEPIYNGHNGENYSYRSVYQCNAGPRLI 2466
Db      2185 LVSHVRECLANGLMSSGSETRCLAGHGPSPIYNGHISGDFSYRDTVYQCNPFRLV 2244
Qy      2467 GMSVRIQQODHMGKTPFCVPTICGHPGNVNGLTQGNOPNLNDVYKFCNNGVMAEGA 2526
Db      2245 GTSVRIQLQDHKMSGQIPVCPTICGHPGNHGLTNGSEFNLDVNFCTHGTGLQGA 2304
Qy      2527 ARSCCLASGQWSDMLPTCRILINCTDPGHQENSVAQVHASGPHRESFGITVSYRQNHGYL 2586
Db      2305 SRACRSNGQWSSPLPICRVVNCSDPGFVENAVHGOQNFPESEHYGTSVMYHCKKGFYL 2364
Qy      2587 LGTPVLSCQGDGTWDRPPOCLVSCGHPSPHSQMSGSYTVGANVRSCTGKRTLVG 2646
Db      2365 LGSSALTCMASGLMDRLPKCLALSCGHPVANAVALTGLFTFGAIVQYSCKGGQILTG 2424
Qy      2647 NSTRMGLDGHMTGSLPHSCGTSVGVCGDPIPAHGIRLDSDPDPTVMSFCEAGHVLIR 2706
Db      2425 NSTRVCOEDSHWSGSLPHSCGNSPFGCDPOTPAHSGRLDDEFKXTSLAFSCEMHQLR 2484
Qy      2707 GSSRRTQANGSWSGSPGEGVISCNPRTPSNARVVSGLVYFSSITVECEGYATG 2766
Db      2485 GSARITLVNGSWGVPCVCEAVSCNRPPTNGMILSSDGIIFSSSVIYACWEGYKTSG 2544
Qy      2767 LLSRHCNVNGTWTGSDDECLVINGDPIRANGRLANDPRYKVTYQCVPMMSHR 2826
Db      2545 LMRHCANGTWTGTADCTILISGDDGTLPNGIQRTDPTFNKTVSYQCNFYLMEPPT 2604
Qy      2827 VSVLSTKDRITWGTFRVCKALMCKPPLIPNGKRVSGSDPMWSSVYATACLEGYQLSLPA 2886
Db      2605 SPTRICKDGTWQTRBLCKAVLCSQPPSVPNKGVESDPRMCASISYSCVDSGYQSHSA 2664
Qy      2887 VFTCEGNSWTGELPOCFPVFCGDPGVPBSRGREDGFSYRSVSCHPRLVTVSSPR 2946
Db      2665 ILSCEGRGVKGVPCQLPVFCGDPGVPBGRSLSGKSFYFKSEVFTQCKRPPVLVSSSR 2724
Qy      2947 FCQSDGTWGTQSDICIDPTLTTCADPVPQFGIQQNSQGVQVSTVLFRCQKGYLLQST 3006
Db      2725 TCQADGWSGSIQPTCIDPAHTACPDGTPHFGIQQNSKGEVSTVFFCRCKGYHQQST 2784
Qy      3007 TRTCLPNTLWSTGTPPCVPHHCQPTPTHTANGLALDPSMGYTLI 3052
Db      2785 TRTCLANLWTSGIQTECIPHACQBPPTPAHADVRAIDLPAFGYTLV 2830

```

```

RESULT 15
AAE20787
ID AAE20787 standard; protein; 3069 AA.
XX
AC AAE20787;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human C3b/C4b complement receptor like protein #1.
XX
KW Human; C3b/C4b complement receptor-like molecule; immune system disorder;
KW gene therapy; rheumatoid arthritis; psoriatic arthritis; osteoarthritis;
KW inflammatory arthritis; inflammatory joint disease; Alzheimer's disease;

```

```

KW multiple sclerosis; inflammatory bowel disease; nervous system disorder;
KW transplant rejection; autoimmune disease; ischaemic condition; neotrophic;
KW metabolic disorder; immunomodulatory; neuroprotective; diabetes; lupus;
KW infertility; vasodilator; obesity; cardiac.
XX Homo sapiens.
FH Key Location/Qualifiers
FT Misc-difference 387
FT Misc-difference /label= Unknown
FT Misc-difference 461
FT Misc-difference /label= Unknown
FT Misc-difference 586
FT Misc-difference /label= Unknown
FT Misc-difference 1272
FT Misc-difference /label= Unknown
XX MO200210199-A2.
XX
XX 07-FEB-2002.
XX
XX 24-JUL-2001; 2001WO-US023232.
XX
XX 02-AUG-2000; 2000US-0222504P.
XX 28-NOV-2000; 2000US-00728787.
XX
XX (AMGE-) AMGEN INC.
XX
XX Welcher AA, Elliott GS;
XX
XX WPI; 2002-303934/34.
XX N-PSDB; AAD33318.
XX
XX Nucleic acid encoding a novel C3b/C4b Complement Receptor-like nucleic
XX acid molecule, useful for treating, preventing and diagnosing rheumatoid
XX arthritis, psoriatic arthritis, inflammatory arthritis, and multiple
XX sclerosis.
XX
XX Claim 13; Fig 1; 251pp; English.
XX
XX The invention relates to a nucleic acid encoding a novel C3b/C4b
XX complement receptor (CR)-like nucleic acid molecule. The C3b/C4b CR-like
XX polypeptide and nucleic acid molecules may be used to treat, prevent,
XX ameliorate, diagnose and/or detect diseases such as immune system
XX disorders such as rheumatoid arthritis, psoriatic arthritis, inflammatory
XX arthritis, osteoarthritis, inflammatory joint disease, autoimmune
XX disease, multiple sclerosis, lupus, inflammatory bowel disease,
XX transplant rejection, nervous system disorders (e.g. Alzheimer's
XX disease), ischaemic conditions, metabolic disorders (e.g. obesity and
XX diabetes) and infertility. The invention is useful in gene therapy. The
XX present sequence is human C3b/C4b complement receptor like protein. Note:
XX The present sequence is stated to be the same as that referred to as SEQ
XX ID NO:2 (AAE20900) shown in page 176-189 of the specification. However
XX the sequences differ at position 695
XX
XX Sequence 3069 AA:
XX
XX Query Match 63.0%; Score 10701.5; DB 5; Length 3069;
XX Best Local Similarity 65.7%; Pred. No. 0;
XX Matches 1833; Conservative 419; Mismatches 531; Indels 5; Gaps 1;
Qy 270 APG-----TELEQSCGDPGIPAYGRREGSFRHHGDTLKEFCQAPAFELVQKAITCOKN 324
Db 17 SPGRVAVYQVEIKEGGCGDPIPAVKRTGSSFLHGDLTLEFCRAAFELVERVITCOQNN 76
Qy 325 QMSAKKFCVNSCFPNTSPSGVTLSPNYPEDYGNHLHCWLIAPDESRIHLAFNDIDV 384
Db 77 QMSGNKSCVNSCFPNTSPSGVTLSPNYPEDYGNHLHCWLIAPDESRIHLAFNDIDV 136
Qy 385 EPQDFLIVIKGATVAPVYGTSPGNTLPSISGSHVATLEPQTHSTKRGNFIFTT 444
Db 137 EPQDFLAVAVDDGSDITLVGTSPGNEVPSQLASSGHVATLEPQTHSTKRGNFIFTT 196

```

QY 445 FRHNECDPGVNVGKAFGBSLOLQSSISFLCDEGFLTQOSENTITCVLKEGSVNMSAV 504
Db 197 FGONECHDPCIPIGRBRFGDRFLIGSSVSFHCDDGFKTQGSSEITITLLODGNVMSSTV 256
QY 505 LRCEAPCGHLLTSBSGTLSPGBMGFYKDALSCAMVLEAOPGYKIKITFDRFKTEVNDT 564
Db 257 PRCAFPCCGHLLTASSGVLTPGMPGYKXDSLHCWMIIEAKGHSIKITFDRFOCEVNDT 316
QY 565 LEVRDGTYSAPLIGVHGTQVPOPLISTSNVYLSTDSKSHDITFOLRYETITLQSD 624
Db 317 LEVNDGASSSPLIGEHGTQAPQPLISTGNFMVLTFTDNRSSISIFLLHYEVSITLSD 376
QY 625 HCLDPGIPVNGORHGNDFYVCAVLTFSQDSGYTSDSEPLECEBNFQMSRLPSCBALCG 684
Db 377 SCIDPGIPVNXHNRHGGDFGIRSTYTFSCDCPGYITSDDEPLVCENHOMHNLPCDMLCG 436
QY 685 GFIOGSSGTLSPGPDYFNNLNCWIIETSHGKVFFTHPHELSGHDYLLITENG 744
Db 437 GYIOGKGTVLSPEFPDFYNSLNXWTIEVSHGKVQMI FHTHLESSHDIITLITEDGS 496
QY 745 FTQPLRLTGSRLPAPISAGLYGNFTQVNFISDPSNYSBGNITTEBYDLEPCEBEVP 804
Db 497 FSEBVALTGSVLEPHITKAGLFGNFTQLRPISDPSISYBGFNTTFSEBYDLEPCDDGPV 556
QY 805 AYAIRKGLQGVGDTLTFSCPGYRLGRTARITCLGRRRLMSPLPRCAECGNSYTG 864
Db 557 AFBRICFHFVGDSLTFSCTGLYRLBGAKKLTLGGGRKRWMSAPLPRCAECASVXGN 616
QY 865 QGTLISNPNVNNNHNCIYSIOTOPKGIOLKARAFELSEGDVLYVNDGNNSARLLG 924
Db 617 EGTLSNPNPNVNNHNCIYKIETBAGKHILRTSFQLEGGDTLKVYDOKSSSPLG 676
QY 925 VFSHSEMMGTVLTSSSLMDFTTDABNTSKGELHFSFELIKCEDPGTKPKGYVND 984
Db 677 TPTNELLGLILNSTSHMLLEFNTNSDTPDQGFQLYTSPDLVCKCDPGIPNNGYIRD 736
QY 985 EGHFAGSSVSFSCPGYSLRGSEBELLCSGERATWMDPLPTCVACCGGYRAGEVGSVLS 1044
Db 737 EGHFTDVLVLSYSCNPGYAMGNSNTLTLCSGDRVWDKPLPSCLAECCGQIHAATSGILS 796
QY 1045 PGYAPARYEHLNLCMTIIEBAGCTIGLFLVFTDEEVDVLRIMDGVESGVLLKELSGP 1104
Db 797 PGYAPARDNNLHCWIIIEADPGKITSIHFIVFTDEMAHDLKVMDDGVDSOILLKEMSGS 856
QY 1105 ALPEDLHSTRSVLTOSTDFTFSTKQGFALIOFSVSTATSCNDPGIPONGSRGCSWAGD 1164
Db 857 ALPEDLHSTRSVLTOSTDFTFSTKQGFALIOFSVSTATSCNDPGIPONGSRGCSWAGD 916
QY 1165 STVFOCDPGYALOGSARISCVKIENRFPMWOPPTCIAPCGGDLTGSBGVLSPNYEPY 1224
Db 917 TMTFOCDPGYALOGSARISCVKIENRFPMWOPPTCIAPCGGDLTGSBGVLSPNYEPY 976
QY 1225 PRGKCEMKTYSBPDYIALVFNIEPEGYDPLHIDGRDLSPLIGSFYGSOLPGRIE 1284
Db 977 PRGKCEMKTYSBPDYIALVFNIEPEGYDPLHIDGRDLSPLIGSFYGSOLPGRIE 1036
QY 1285 SSSNSLFLAFRSDASVSNAGFVIDYENPRESCTDPGSIKXGTRVSGDLKLGSSVTTYCH 1344
Db 1037 SSSNSLFLAFRSDASVSGFAIEFKKPREACDPMGNMGTVGDTDFLKGSTIITYQCD 1096
QY 1345 GGYVEBGTSLTCLSGDQKPVMMNPRPVCTAPCGGQYVSDGVLSPNYPOUNTISOIC 1404
Db 1097 GGYVEBGTSLTCLSGDQKPVMMNPRPVCTAPCGGQYVSDGVLSPNYPOUNTISOIC 1156
QY 1405 LYFVTEKDYVVFQOFAFHTALNDVVEVHDGSHQSHRLSSLSGSHTGESLPLATSNV 1464
Db 1157 LYFVTEKDYVVFQOFAFHTALNDVVEVHDGSHQSHRLSSLSGSHTGESLPLATSNV 1216
QY 1465 LKRSAGLAPARGHFFYQAVPRTSATQCSVPEPRYKGLKSGDFYVGAIVRECKSGY 1524
Db 1217 LKRSAGLAPARGHFFYQAVPRTSATQCSVPEPRYKGLKSGDFYVGAIVRECKSGY 1276
QY 1525 ALQSPREIBCLPVFGALAQNMVSAPTCVVPCGGLTERRGTLISBPPEPYLNSLNCVMK 1584
Db 1277 LLQGSTALHQSPVNLMAQNMNDITPSCVVPSCGNFTQORRTILSPGPEBYGNLNCVMK 1336
QY 1585 IYVPEGAGIOIYQVSVFTBQNMDSLEVFGADNTVMLSGSGTTPALINSTNSOLYLH 1644
Db 1337 IYVTEGSGIOIYQVSVFTBQNMDSLEVFGADNTVMLSGSGTTPALINSTNSOLYLH 1396
QY 1645 FVSDISVSAAGPFLKXTVGLSCPEPVAVSNVKTGERLVNDVVSFOCEPYALQGH 1704
Db 1397 FVSDISVSAAGPFLKXTVGLSCPEPVAVSNVKTGERLVNDVVSFOCEPYALQGH 1456
QY 1705 HISCMPTVRRMNYPPPLCIAQCGTVEEMEGVILSPGPBNYPSNMDCSMKIALPVFG 1764
Db 1457 HISCMPTVRRMNYPPPLCIAQCGTVEEMEGVILSPGPBNYPSNMDCSMKIALPVFG 1516
QY 1765 AHIOPLNFSTEPHNDYIETRNQYETSRMMGRPSGSELPSLSTSHETTVYHSDHSON 1824
Db 1517 AHIOPLNFSTEPHNDYIETRNQYETSRMMGRPSGSELPSLSTSHETTVYHSDHSON 1576
QY 1825 RRGFKLEYQAYELOECPDEPFPANGIVRAGVNVGGSVTFECLPGYOLTGHVPLTQOHT 1884
Db 1577 RRGFKLEYQAYELOECPDEPFPANGIVRAGVNVGGSVTFECLPGYOLTGHVPLTQOHT 1636
QY 1885 NRMNDHPLPKCEVPCGANTISSNGTVYSPGFPSPYSSODCWMLITVPLIGHVRLNLSL 1944
Db 1637 NRMNDHPLPKCEVPCGANTISSNGTVYSPGFPSPYSSODCWMLITVPLIGHVRLNLSL 1696
QY 1945 QTEPSGDTITWDGPOQTAARLGVFTRSNAKKTVOSSNSOVLKFRHDAATGGLFALAFS 2004
Db 1697 QTEPSGDTITWDGPOQTAARLGVFTRSNAKKTVOSSNSOVLKFRHDAATGGLFALAFS 1756
QY 2005 AYPLTCRPPRTLLPNEVLTENEFPNIGDIVRRCILPGFLVGNELITCLGTVLOPEGP 2064
Db 1757 AYPLTCRPPRTLLPNEVLTENEFPNIGDIVRRCILPGFLVGNELITCLGTVLOPEGP 1816
QY 2065 PPIICEVHCPTEBELTSTGVILISQSYPGSPQOTCSMTLVRBPDVNI SLTVEYFLSEKQ 2124
Db 1817 PPIICEVHCPTEBELTSTGVILISQSYPGSPQOTCSMTLVRBPDVNI SLTVEYFLSEKQ 1876
QY 2125 YDEFEITFDGSGSPLIKALSGNYSAPLYTSSNSVYLRMSDDAHVNRKGFIRISAPY 2184
Db 1877 YDEFEITFDGSGSPLIKALSGNYSAPLYTSSNSVYLRMSDDAHVNRKGFIRISAPY 1936
QY 2185 CSLPRALHGFITIGOSTPGCSIHFGCNAVYLVGSHMALCIRHPGYVLMSEALPLCQ 2244
Db 1937 CSLPRALHGFITIGOSTPGCSIHFGCNAVYLVGSHMALCIRHPGYVLMSEALPLCQ 1996
QY 2245 ALSGCLPEAPKGMVFGKETTGTAKVYSCSEGYHLQAGAEATACIDTGLMGNRPVPO 2304
Db 1997 ALSGCLPEAPKGMVFGKETTGTAKVYSCSEGYHLQAGAEATACIDTGLMGNRPVPO 2056
QY 2305 CVPTCEPVSSISVEHGRWRLIETQYQFOALMLICDPGYTGTQORVIRCOANGKWSL 2364
Db 2057 CVPTCEPVSSISVEHGRWRLIETQYQFOALMLICDPGYTGTQORVIRCOANGKWSL 2116
QY 2365 DSTPTCRITISGSLPLIPRNHRIGTLSVYCATIIFGNSGYTLVGSVRECMANGMLSGS 2424
Db 2117 DSTPTCRITISGSLPLIPRNHRIGTLSVYCATIIFGNSGYTLVGSVRECMANGMLSGS 2176
QY 2425 EYRCLAGHCTPEPIVNGHINGENYSYRGSVVYOCNAGFLIGMSVARI COQDHHWSGKT 2484
Db 2177 EYRCLAGHCTPEPIVNGHINGENYSYRGSVVYOCNAGFLIGMSVARI COQDHHWSGKT 2236
QY 2485 FCVPITGHPGNPNVNLTOGQFNLDVYKVCNPGYMAEGAARSOCLASGOWSDMLPTC 2544
Db 2237 FCVPITGHPGNPNVNLTOGQFNLDVYKVCNPGYMAEGAARSOCLASGOWSDMLPTC 2296
QY 2545 RIINCDPHGQENSVPQVHASGPHRPSFGTTSYRCNCHGYLLIGTPLYLSCGGSGTMDPR 2604
Db 2297 RIINCDPHGQENSVPQVHASGPHRPSFGTTSYRCNCHGYLLIGTPLYLSCGGSGTMDPR 2356
QY 2605 POCLIVSCGHPGSPPHSOMSGSDYTVGAVRYSICIGRTLVGNTMCGIDGMWTSGLPH 2664
Db 2356 POCLIVSCGHPGSPPHSOMSGSDYTVGAVRYSICIGRTLVGNTMCGIDGMWTSGLPH 2664

Db 2357 PKCLAI SCGHPV PANAVLTNGELFTYGA VVHYSCRGSESLI GNDTRVQEDSHWSGALPH 2416

QY 2665 CSGTSVGVCGDPGIPAHGIRLGDSPDPTWRFSCGAGHYLRGSSERTCOANGSWGSSOP 2724

Db 2417 CTGNMPEFCGDPGTPAHOSRLGDDPKTKSLRFSCEMHQLRGSPERTCLNLSWSGLQP 2476

QY 2725 ECGVISCNPGTPSNARVVFSDGLVFSSIVYECREGYATGLLSRHCSVNGTWGSDPE 2784

Db 2477 VCEAVSCGNPGTPTNGMIVSSDGLFSSSVIYACMEGYKTSGLMTRHCTANGTWGTAPD 2536

QY 2785 CLVINCDBGIPANGRLGNDFRNKTYYTQCVPGYMMESHRSVLSCTKORTNGTKPY 2844

Db 2537 CTIISCGDPGTLANGIOFGTDFTFNKTVSYQCNPGYMEAVTSATIRCTKDXGRNPSKPV 2596

QY 2845 CKALMCKPPLIPNGKTVVSGDFMWSSTVYACLEGYQLSLPAYPTCEGNSWTGELPOCF 2904

Db 2597 CKAVLCQPPPVQNGTVEGSDFRWSSISYSCMDGYQLSHALISCEGRGVWKGEIPQCL 2656

QY 2905 PVFCGDPGVBSRGREDRGSYSRSVSFSCHPPLVLVGSPPRFQSDGTWSTGTOPSCIDP 2964

Db 2657 PVFCGDPGIPAEGRLSGKSFYKSEVFFQCKSPILVSSRRVQADGTWSGIQPTCIDP 2716

QY 2965 TLTTCADPBGVQFQIQQNNSQGYQVSTVLRCCQKGYLLQGSTTRTCLPNLTWSTGTPPDCV 3024

Db 2717 AHNCTCPDPGTPHFGIQQNSSRGYEGSVTFPRCRKGYHIQGSTTRTCLANLWMSGIOJTECI 2776

QY 3025 PHHCROPEPTTHANVGALDLPMSMGYTLI 3052

Db 2777 PHACRQPEPTPAHADVRALDLPFTGYTLV 2804

Search completed: October 18, 2004, 10:21:45
Job time : 125 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model1

Run on: October 18, 2004, 10:04:11 : Search time 35 Seconds

(without alignments)
8533.052 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPPPALLPCLSLSDCC.....RSGPVGDPSTLPGRSHPKP 3104

Scoring table:

Gapop 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

1: PIR.79:*
2: PIR1:*
3: PIR3:*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1555	9.2	2489	2	173012 complement C3b/C4b
2	1473	8.7	3623	2	109456 intrinsic factor-B
3	1404.5	8.3	3623	2	108618 intrinsic factor-B
4	1390	8.2	2014	2	136936 complement recepto
5	906.5	5.3	1025	1	A43526 complement C3d/Bps
6	904	5.3	1091	1	PI0009 complement C3d/Bps
7	763.5	4.5	977	2	152657 seizure-related pr
8	755	4.4	991	2	149540 procollagen C-endo
9	746	4.4	986	1	B58768 complement factor
10	721	4.2	1231	1	NBHUH complement factor
11	685.5	4.0	3871	2	T22812 hypotethical prote
12	677	4.0	676	2	A45900 complement C3b rec
13	676.5	4.0	1234	1	NEMSH complement factor
14	657.5	3.8	830	2	A30359 P-selectin precurs
15	640.5	3.8	1070	2	T31069 CRP-ductin-BMP-1 like
16	626.5	3.7	2083	2	T42721 P-selectin-alpha p
17	605	3.6	768	2	A42755 P-selectin precurs
18	597	3.5	768	2	I53821 P-selectin - rat
19	577.5	3.4	1057	1	A39288 dorsal-ventral pat
20	576	3.4	1464	2	S58954 development protei
21	556	3.3	1797	2	T21889 hypotethical prote
22	547	3.2	1805	2	T21888 hypotethical prote
23	545	3.2	1524	2	T30337 polyprotein - Afri
24	542.5	3.2	1053	2	S46159 probable complemen
25	526	3.1	610	2	A35046 E-selectin precurs
26	514	3.0	767	2	T30018 hypotethical prote
27	505.5	3.0	597	1	S53711 C4BP alpha chain p
28	504	3.0	707	2	JC2218 procollagen C-endo
29	501.5	3.0	1827	2	T34288 hypotethical prote

30	498.5	2.9	730	1	BNHUI	procollagen C-endo
31	498.5	2.9	823	1	A58788	procollagen C-endo
32	491.5	2.9	597	1	NBHUC4	C4b-binding protei
33	490	2.9	808	2	D35069	complement factor
34	489.5	2.9	560	2	T16833	hypotethical prote
35	481.5	2.8	612	2	B42755	E-selectin precurs
36	477	2.8	610	1	I46001	C4b-binding protei
37	468.5	2.8	37	1	S57953	C4BP protein alpha
38	466	2.7	482	2	A34924	complement C3b/C4b
39	463	2.7	669	2	S65551	factor H - bovine
40	461.5	2.7	646	2	JN0473	P-selectin precurs
41	458.5	2.7	1290	2	A57190	ebnerin precursor
42	452.5	2.7	2043	2	T18524	scavenger receptor
43	450.5	2.7	579	2	A56740	sperm-egg recognit
44	442	2.6	1019	2	A38738	coagulation factor
45	404	2.4	551	2	I46709	endothelial leukoc

ALIGNMENTS

RESULT 1
173012
complement C3b/C4b receptor, membrane-bound form precursor - human
N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); surfa
N:Contains: complement C3b/C4b receptor, secreted form
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence revision 24-Nov-1999 #text change 09-Jul-2004
C:Accession: 173012; I56203; A47602; S03843; A28507; A24748; B24748; C24748
R:VIK, D.P.; Wong, W.W.
J:Immunol. 151, 6214-6224, 1993
A:Title: Structure of the gene for the F allele of complement receptor type 1 and seque
A:Reference number: I56203; MUID:94065175; PMID:8245463
A:Accession: 173012
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-683; 'X', 685-894; 'A', 896-1000; 1451-1471; 'X', 1473-2489 <VIK1>
A:Cross-references: GB:LI7418; NID:G306678; PIDN:AB60694.1; PID:G306680
R:Mong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.; W
J:Exp. Med. 169, 847-863, 1989
A:Title: Structure of the human CR1 gene. Molecular basis of the structural and quant
A:Reference number: A47602; MUID:89176869; PMID:2564414
A:Accession: A47602
A:Molecule type: DNA
A:Residues: 1-41 <MON>
A:Cross-references: GB:X14893
R:Hourcade, D.; Meisner, D.R.; Atkinson, J.P.; Holers, V.M.
J:Exp. Med. 168, 1255-1270, 1988
A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b rec
type 1.
A:Reference number: S03291; MUID:89010527; PMID:2971757
A:Accession: S03291
A:Molecule type: mRNA
A:Residues: 26-584 <HOU>
A:Cross-references: EMBL:X14362; NID:G30197; PIDN:CAA2541.1; PID:G736240
A:Experimental source: Clone CR1-4
R:Klickstein, L.B.; Barrow, T.O.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.
J:Exp. Med. 168, 1699-1717, 1988
A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b
A:Reference number: S03843; MUID:89035992; PMID:2972794
A:Accession: S03843
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-894; 'A', 896-1000; 1451-2064; 'T', 2066-2276; 'P', 2278-2299; 'H', 2301-2325; 'T',
A:Cross-references: EMBL:Y00816; NID:G30185; PIDN:CAA68755.1; PID:G30186
R:Klickstein, L.B.; Wong, W.W.; Smith, J.A.; Weis, J.H.; Wilson, J.G.; Fearon, D.T.
J:Exp. Med. 165, 1095-1112, 1987
A:Title: Human C3b/C4b receptor (CR1). Demonstration of long homologous repeating domain

A:Reference number: A28507; MUID:87168191; PMID:2951479
A:Accession: A28507
A:Molecule type: mRNA
A:Residues: 955-1221, 'FV', 1224-2064, 'I', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2
A:Cross-references: GB:X05305; NID:930196; PIDN:CA28953.1; PID:9809019
R:Wong, W.W.; Klicstein, L.B.; Smith, J.A.; Weis, J.H.; Fearon, D.T.
Proc. Natl. Acad. Sci. U.S.A. 82, 7711-7715, 1985
A:Title: Identification of a partial cDNA clone for the human receptor for complement Ff
A:Reference number: A94073; MUID:86067975; PMID:2933745
A:Accession: A24748
A:Molecule type: mRNA
A:Residues: 311-333/729-745/831-845 <MO2>
A:Cross-references: GB:M11569; NID:9180991; PIDN:AAA52297.1; PID:G180995; GB:M11617; NID
C:Genetics:
A:Gene: GDB:CR1; CD35
A:Cross-references: GDB:119800; OMIM:120620
A:Map position: 1q32-q33
A:Interon: 41/1; 101/1; 134/2; 163/1; 296/1; 355/1; 389/2; 418/1; 491/1; 551/1; 584/2; 6
; 1464/2; 1513/1; 1646/1; 1705/1; 1739/2; 1768/1; 1844/1; 1904/1; 1937/2; 1966/1; 2099/1
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-41/Domain: signal sequence #status predicted <SIG>
F:42-2489/Product: complement C3b/C4b receptor, membrane-bound form #status predicted <M
F:42-584/Product: complement C3b/C4b receptor, secreted form #status predicted <MATS>
F:43-95/Domain: complement factor H repeat homology <FH01>
F:104-161/Domain: complement factor H repeat homology <FH02>
F:166-232/Domain: complement factor H repeat homology <FH03>
F:228-233/Domain: complement factor H repeat homology <FH04>
F:297-353/Domain: complement factor H repeat homology <FH05>
F:358-416/Domain: complement factor H repeat homology <FH06>
F:421-487/Domain: complement factor H repeat homology <FH07>
F:493-549/Domain: complement factor H repeat homology <FH08>
F:554-611/Domain: complement factor H repeat homology <FH09>
F:616-682/Domain: complement factor H repeat homology <FH10>
F:688-743/Domain: complement factor H repeat homology <FH11>
F:747-803/Domain: complement factor H repeat homology <FH12>
F:808-866/Domain: complement factor H repeat homology <FH13>
F:871-937/Domain: complement factor H repeat homology <FH14>
F:943-999/Domain: complement factor H repeat homology <FH15>
F:1004-1061/Domain: complement factor H repeat homology <FH16>
F:1066-1133/Domain: complement factor H repeat homology <FH17>
F:1138-1193/Domain: complement factor H repeat homology <FH18>
F:1197-1253/Domain: complement factor H repeat homology <FH19>
F:1258-1316/Domain: complement factor H repeat homology <FH20>
F:1321-1387/Domain: complement factor H repeat homology <FH21>
F:1393-1449/Domain: complement factor H repeat homology <FH22>
F:1454-1511/Domain: complement factor H repeat homology <FH23>
F:1516-1582/Domain: complement factor H repeat homology <FH24>
F:1588-1643/Domain: complement factor H repeat homology <FH25>
F:1647-1703/Domain: complement factor H repeat homology <FH26>
F:1708-1766/Domain: complement factor H repeat homology <FH27>
F:1771-1837/Domain: complement factor H repeat homology <FH28>
F:1846-1902/Domain: complement factor H repeat homology <FH29>
F:1907-1964/Domain: complement factor H repeat homology <FH30>
F:1969-2035/Domain: complement factor H repeat homology <FH31>
F:2041-2096/Domain: complement factor H repeat homology <FH32>
F:2100-2156/Domain: complement factor H repeat homology <FH33>
F:2161-2219/Domain: complement factor H repeat homology <FH34>
F:2224-2290/Domain: complement factor H repeat homology <FH35>
F:2298-2354/Domain: complement factor H repeat homology <FH36>
F:2359-2415/Domain: complement factor H repeat homology <FH37>

Query Match 9.2%; Score 1555; DB 2; Length 2489;
Best Local Similarity 20.9%; Pred. No. 1.6e-79;
Matches 679; Conservative 332; Mismatches 1021; Indels 1224; Gaps 142;

QY 58 KTSVLQYGVSGQHNMCPDGPPIPERGK-RLGSDFRIGSVQFCNENGLQSSKRTCKM 116
DB 88 KNSVWGAKDRCKRCKRNPDPVNGMVHVKIGIQGSIKYSCTKGYRLISSSATCT- 146
QY 117 VSDMFAWSDHRCVCAKMCDAHLRGPSSGIIITSPNPIQYDNNNAHCWVIT-ALNPSKVI 175
DB 147 ISGDVTYIMNETPIGCRIC-----GLPPIITNGDPISTNRNENHAGSVVTRCNPGSGG 201

QY 176 KLAFFEPDLERGVDTLVEDGGDDQKTVLYMSQACSDSPHTGSRIPESNSGDIWQ 235
DB 202 KRYFE-----LVGEPS-----ICTSN-----DQVG-IV-- 225
QY 236 KMTVLERICDISSDARSVKRSPKTSNAVELVAPGTIEQSGSDP-----GIPIYGR 291
DB 226 -----SGAPQCIIT-----NKCTPPVENGLVSDNR 253
QY 292 EGSRFPHSGDLKRECQPAFELVQKAITQKNNQSAKKPGCVFSGFNFTSPSGVLS 351
DB 254 --SLFSLNVEVEFRCPGPFVMKGPRRVKQALNKVEPELPSGSRVC-----QPPDVLHAE 307
QY 352 NYDEVDGNHLHCWLLIAPESRIHAFPDIDVEQFDPVLKDGATAPALGTFSGNQ 411
DB 308 RTQRDKNF-----SPGEVYS-----CEPDYD-----LRGAASMRCTPGDMS-- 347
QY 412 LPSSTTSGHVARLEFQDHTSGKRGFNITFTFRHNECPD-PGVPVNGK-RFGSLQIG 469
DB 348 -PAA-----PTCEVKSDDPMGQLNGRVLPVNLQIG 379
QY 470 SSISFLCDGFLGTGSETTITVLKGSVWNSAVLRCEAPCGHLTSPSGTILSPGMPG 529
DB 380 AKVDFVCEBGF-QLKGSASAYCVLAGMESLMSVPCG-----QIFGSPVPIWG-- 430
QY 530 FYKDALSCMWLEAQPGYPIKITPRFTEVYD-----TLFVR 568
DB 431 -----RHTGKPLEVF--PGKAVVYTCDDPHDRGTSFDLIGESTIRCTSPQ 475
QY 569 DGRYTS--APLIGVHGTQVPOFLISTNSNYLLFSTKSHSD-IGFOLRYE----- 617
DB 476 GNGVWSPAPRCGILGHQAP-----DHFPLAKLQTQMSDFPIGSLKTECPREY 529
QY 618 ---TTLQSDHCLD-----PGIPVNGQRHG-NDYVAGALVTFSCDSG 655
DB 530 RPSFIT-----CLDWLVWSSPDVCKRKSCTPPDPVNGMVHVTIDVGSIRYNSCTTG 584
QY 656 YTLISGEPLECEPN---PQMSRALPSCEAL-CGGFIQSSGTTILSPGPDYPPNNLNTW 711
DB 585 HHLIGHSABCLISGNAAMHSTKPTICGRICG----- 617
QY 712 ILETSHGKGVFTFTFTHLESQHDVLLITENGSTFQPLRLQUTGSRLPAPISAGLYGNFTA 771
DB 618 -----LPTIANG----- 625
QY 772 QVRFTISDF-SMSYEGFNITPSEYDLFCEBEVEVPAYSIRKGLQGVGCTLLPSCPG-- 827
DB 626 ---DFTSTNRENPH-----YGSVVYVYRCNPGSGG 651
QY 828 ---YRLBGTARITCLGGR--LWSSPLPRCAEGNSVTGTQTLISP--FPVNNYNN 880
DB 652 RKVPELVBEPIYCTNSNDQVIGMSGAPQCIIT-----PNKCTPPVENZG 696
QY 881 HECTIYSIQTPQKGIQLKARAFELSEGDVLKVYDGNNSNARLLGVFSHEMKVGTLNSTS 940
DB 697 -----ILVSDNRS-----LFSINEV----- 711
QY 941 SSIWLPDITDAENSTKGFELHFFSPELICKEDPQTPKFGYVHDEGHAGSSVSPSCDPG 1000
DB 712 -----VEFRQCPG 719
QY 1001 YSLNGSEBLLCLSGERTWRPLPTCVAECGTVRGEVSGQVLSPPGAPAPAEHNLNCTWT 1060
DB 720 FVMGSPRRVKQALNK--WEPELPSGSRVC-----QPP----- 751
QY 1061 IEADAGCTIGHLVPTDEEVHVLRIWGDVESGVLKELGPALEPKDLHSTFNSVVLQ 1120
DB 752 -----DVLIH----- 755
QY 1121 FSTDFPFSKGFALQFVSSTATSCNDPGIIPONGSRSDSWEAGDSTYFQCDPGALOGSA 1180
DB 756 -----AERTQRDKDNFSPGGEVFEVPSCEPGVLDLGA 786

QY 1181 EISCXKIENREFWOPSPPTCIAPCGDLTGP--SGVILSPNYPEPPKCEKDWKVTSP 1238
 Db 787 SMRTCTPGD---NSPAPCTEVKSCDDPMGLANGRVLP----- 823
 QY 1239 DYVIALVFNILNLEPGYDFLHIYDGRDLSPLIGSFYGSQLPGRIBSSNSLFLA----- 1293
 Db 824 -----VNLQGAKVDFVCD-----EGFOLKG---SSASVCVLAGEMSL 858
 QY 1294 PRSDASVSMAGFVLDYIENPRESECPDGSINQTRVSGDLK---GSSVTTYGH----- 1344
 Db 859 WNSVSPVCEQIF-----CPSPVINGRHRTGKPLVFPFGKTVNTYTCDPHPDRG 907
 QY 1345 GGYVEGTSTLSCILGPDGKPVMMNPRVCTAPCGGVGSDGVLSPNYQNTSGQIC 1404
 Db 908 TSPILIGESTIRCTSDPDGNGWSSPAPRC-----GIL----- 940
 QY 1405 LYFVTPKDVVFGQAFHTALNDVVEHDHSGQ--HSRLLSLSGSHGESLPLATSN 1462
 Db 941 -----GHCOAPDHFLFAKLKTQTNASDPFGTS- 968
 QY 1463 QVILKFAKGLAPARGF-----HFVQAVPR--TSAQCSSVPEPRYK-RLGSDPSVG 1513
 Db 969 ---LKYECPREYGRPFSTICTLDNLWSS--PKDVCKRKSCKTPDPVNGVHVITDIQVG 1024
 QY 1514 AIVAFECNSGYALQSPREIECLPVGALQNNVSAPTCV--VPCGANTERRGTLSPGF 1572
 Db 1025 SRIVYSCCTGHRILGHSAECI--LSGNAAMHSTKPIQRIPLGLPIIANGDITSTRE 1083
 QY 1573 EPLVNSLNCWKIIVPEBAGIQIOVVSFVTEQNNDSLLEVFDADNTVTMLGSFSGTTPVA 1632
 Db 1084 NFHYGS-----VVTYRCNPGSGRKVFE-----LVGEPS----- 1112
 QY 1633 LUNSTNQYLHFFYSDISVSAAGHLEKYTGGLSCPRP--AVPS-----NGVKTGE 1682
 Db 1113 -IYCTSD-----DOYIWSGAPPOCIIPKCTPPEVNGELIWD 1151
 QY 1683 R---YLVDVVSFOCEPGYALQGHANHSCMGTVARMYRPPLCIAOCGVEEMEGYL 1739
 Db 1152 NRSILFSLNVEVEFKQPGFVKKGRVYCO--ALKMPELPSGRVC----- 1197
 QY 1740 SPGEPPGNVPSNMDCKWIALFVGFGAMIQFLNFSTEPNHDYIEIRNGEYETSRMNGRESG 1799
 Db 1198 -----QPPPDVL----- 1204
 QY 1800 SELPSSLLSTSHETTVYTHSDHSQNRPGKLEYQAYLEQECRDEBPANGLVRGAGYVG 1859
 Db 1205 -----HAERTOR-----DKDNF-----SPG 1219
 QY 1860 QSVTFECLPGYQLTGHVPLTCQHGNNRMWHDRLPKCEVPCGANTSS--NGTVVSPGPPS 1917
 Db 1220 QEVYYSCEPGYDLKGASMRCT--POGDMSPAPRCVKSDDPMGLANGRVLP----- 1273
 QY 1918 PYSSSODCVMLITVPIGHVNLNLSLQTEPSGDFITIMDGPQOTAPRLGFTSRMAKT 1977
 Db 1274 -----VNLQGAKV-----DFVCD--EGFOLKG----- 1294
 QY 1978 VOSSNOVLKFPHDAATGIFAFAFAYPLTK---CPPTILPNAEVTNEN--EENIGD 2033
 Db 1295 ---SSASVCVL-----AGMESLWNSVVPVCEQIFCPSPVINGRHRTGKPLVFPFGK 1344
 QY 2034 IVRFRCLP-----GFTLVGNEILTKL-----GTYLQFEGBPPICEV--HCPINELLTD 2080
 Db 1345 AVNTTCDBHPDRGHSFDLIGESTIRCTSDPDGNGW---SSPAPRCGLIGHCOA----- 1395
 QY 2081 STGVILSGSYPSGYPOFQTCMVLVRVEBDYNI SLTVEYFLESEKQYDEFEIFDGSQGPL 2140
 Db 1396 -----PD-----HFLPAKKTQTNADPPIGSLK 1420
 QY 2141 LKALSGNYSAPLITYSSNSYVLKMS--DHAYNKGFIRISAPYCSIPRAPLHGFIIGQ 2199
 Db 1421 YECPEBYVYGRPFSTICTDNLV---WSPKDVCKRS-----CKTPRPVNGVHV 1468
 QY 2200 TSTQPGSIHFGCNAGYRLVGHSMALCTRHPOGYHLMSEALPGLQALSGCLPBAPKQMV 2259

Db 1469 TDQVGRINVSCTTGHRILGHSSAECILSGNTAH--MSTKPIQRIPLCGLPPIIANGDF 1527
 QY 2260 FG---KEYTVGTAIVSCSSEGH---LQGAETAECL-----DTGLMS-----NR 2299
 Db 1528 ISTRNENHYGVSVTYVCNLGSRGRKRVFELVGEPSIYCTSDNDQVIGMGPAPOCIIIPNK 1587
 QY 2300 NVPP-----OCVP-----VTC-----PDVSSIS--- 2317
 Db 1588 CTPPVNENGLIVSDNRSLSFLNVEVEPCQGFYMKGRPRVYKCOALMKMBELPSGRVC 1647
 QY 2318 -----VEHGRWRLIFETOYFOAQLMLICDPGYVYTCQAVYRCQANGKWSLIGSTPCR 2372
 Db 1648 QPPEILHGHETBHQNFSPQGEVYFSCBEGYDLRGAALHCTPGQDWS--PEARCAV 1705
 QY 2373 ISC---GELPIPNHGRIGTLVS--YATAI FSCNSGYTLVGSVRECMANG---LWSG 2424
 Db 1706 KSCDDFLGOL---PHGVLPPLNLQGAKVSFVDEBGFRLKSSVSHCVLVGMRSLWNS 1762
 QY 2425 EVRLAGHGTBPRTVNGHNGE---NYSYRGSVYQCN-----AGFRLIGMSVRIQOQ 2475
 Db 1763 VPVCEHIFCPNPAILNGRHRTGPSGDI PYGKEIYTCDPHPDRGMTFNLIGESTICTS 1822
 QY 2476 DHH-----MSGKTPFC--VPITCGH-----PGNPVGLTQGNFNLNDVYKFCVNP 2519
 Db 1823 DPHNGVWSSPAPRCCLSVAGHCKTPEQPPASPTIPIINF-----EFPVGLSLNYECRP 1878
 QY 2520 GYMAEARSOCCLASGOWSMDLPTCRILINCTDPHQENSVROVHASCPHRSFCTTVSYR 2579
 Db 1879 GYFGKMFPSIS--CLENLWSSVEDNCRKRKSGPPEPPNGVHVIITD---TOFESTVANS 1933
 QY 2580 CNHGFYLLGTPIVUSCOGD---TWDRPPOCLLVSCGHPSPRHSQMSGSYT----- 2629
 Db 1934 CNBSGFRILIGSBSTCLVSGNNVTWDKAPICEIISIC---BPPRTISNGDYISNNRISFH 1989
 QY 2630 VGAVVYRSC-----IGKRTLVGNSSTRMCGLDGHWTSGLPHCSGTSVGCGDPG 2677
 Db 1990 NGTVVITYQCHTGPBGDLFELVGRSLEY--CTSDDQVGVWSSPPRCISGTN--KCTAPE 2045
 QY 2678 IPAHGRIGLDS---FDPGTWRRSCGAYVLGRSSERTCOANGWSGSOBPCGVISCGNP 2734
 Db 2046 V-ENAIIVPGRNRSFSLTEIYFRFCQPFVWVGSHTVQCOYTNGRWKGFKLPHCSRV--COP 2103
 QY 2735 GTPSNARVPS--DGLVFSSSIYECBEGYVANGSLSRHCSVNGTWTGSDPECLVINCQD 2792
 Db 2104 PEILHGHNTLSHONFSPQGEVYFSCPSYDLRKAALHCTPQGDWSPPEARCTVKSDD 2163
 QY 2793 --PGIPANGRLAGNDFYNTKTVTYQCVPGYMMESHVVS--VLSTCKDRTWNGTRPKYKAL 2848
 Db 2164 FLOGLPHGRVULPLNLQGAKVSFVCDGFRILKGRSASHCVLAGMK--ALMNSSVVPVCEQI 2222
 QY 2849 MCKRPPLIPNGKVVGS---DFMGGSVTVAC-----LEGYOLSLPAVFTCEGNG 2894
 Db 2223 FCPNPAILINGRHRTGTFPGDIPYGEKISYACDTHPRDGMTFNLIG--ESSIRCTSDROGNG 2281
 QY 2895 SWTGELOPC---FVFPFGCDGCVBPSRGRREDBRGSF---RSVSFSCBPLVLVVSPPRRC 2948
 Db 2282 VMSSPAPRCCLSVAPACPDPPKIQNGHYIGAHVSLYLPGMTISTICTDGYLVKGFTFC 2341
 QY 2949 QSDTMSGTOPSCIDPLTTCADPGVPOFGIIONN---SOGVGVSTYVLPFCQKQYLLQGS 3005
 Db 2342 TDQISVQDLHCKE---VNCSPF--LPMNGISKELEMKYHYHGDYTLACBODYTLBGS 2397
 QY 3006 TTRTCLPNTLWSGTRP 3021
 Db 2398 PWSQCADDRWD--FP 2411

RESULT 2
 T09456
 intrinsic factor-B12 receptor Cubilin precursor - human
 C/Species: Homo sapiens (man)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T09456
 R:Kozyrak, R.; Kristiansen, M.; Silahcaroglu, A.; Hansen, C.; Jacobsen, C.; Tommerup, N
 Blood 91. 3593-3600, 1998
 A:Title: The human intrinsic factor-vitamin B12 receptor, cubillin: Molecular characteriz
 ion.
 A:Reference number: 216677; MUID:98241400; PMID:9572993
 A:Accession: T09456
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3623 <KOZ>
 A:Cross-references: UNIPROT:O60494; EMBL:AF034611; NID:G3929528; PIDN:AAC82612.1; PID:G3
 C:Genetics:
 A:Map position: 10p12
 C:Superfamily: intrinsic factor-B12 receptor cubillin; EGF homology
 C:Keywords: receptor; vitamin B12 uptake
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-3623/Product: intrinsic factor-B12 receptor #status predicted <MAT>
 F:436-461/Domain: EGF homology <EGF>

Query Match 8.7%; Score 1473; DB 2; Length 3623;
 Best Local Similarity 21.2%; Pred. No. 1-2e-74;
 Matches 795; Conservative 416; Mismatches 1184; Indels 1356; Gaps 174;

113 TCMKSDNFPA-----MSDHRVCRA--RMCDAHLRGPSSGIITSPNPPIQYDNNAHCWVI 165
 446 TCVDGVDSPSCCECTRLMTG--ALCQVPOVCESISGINGSFSYRSPDVGVHDVNCFW 503
 166 ITALNPSTVITLAFEEFLER---GVDTLTGCGGQDQKVTYVMSQNACSSEPTFG 221
 504 IKT-EMGVNLAITTFPLESMNDCPHEFLQV---YDGDSSAFQJGRCCGSLPR--- 555
 222 SRIPESMGDIWRQKMTVLEICRDISSDARSQSVKSPKTSNAVELAVGTETEGSSCG 281
 556 -----ELSSD-----NALYHLXSEHLRNG--- 576
 282 DPGIPAYGRGRSGFHGDTLKEFCOPAFELVQKAITCOKNOWSAKKFCVPSCFENF 341
 577 -----RGFTVR-----METQCP-----ECGGL 594
 342 TSPGVLSPNYPEDYGNHLHCWVLLAPESRIHLAFNDIDVP-----QFDFLVIKDA 397
 595 TGPFGISKSPCPENYPRGDCWIVVTSPLDVTTFRTGLSLHDDCKDKDLLEIRDG 654
 398 TAEAPVLTSGNOLPESITSSGHVAFLEQTHSTGKRGFNITFTFRN--ECPDPGV 456
 655 LYQPLLKGECTTFVSPPLQTTGPFARIHHSQISDQGFHITVLTSPDLACGNYTD 714
 457 VNGKRF-----GDSLQLG--SSISFLCD-----EGF 480
 715 PEGELFLPELSGPTHTRCVVMKQPGEOIQINFTHELOCOSSSONYIEVRDGETL 774
 481 LG-TGSETTTCVLEKGSVW-----NSAVLRC--EAPCGHLTSPGTLTSPGPW 528
 775 LGKVCNGGTSIHKISITNSVWIRFKIDASVEKASPAVVQVACGDLTG--EGVIRSPFP 833
 529 GFYKDALSCMWLEADQGVPIKITPDRK---TEVNYVTLVLRDGRKTSAPLIGVYHGT 584
 834 NVTPGERTCKWTHQPOSQVILNFTVFEIGSSAHCEYVEIGSSSISGSENNKXYCGT 893
 585 QVDPQLISTSNVLYLFTSTDKSHSDIGFOLRYETITLQSDHCLDPEIPIVNGQRHGDFTV 644
 894 DIFSFTSYVNFYLVFVSSSTENHGFMAKF----- 925
 645 GALVTSQDSGYTLSDGEPLCEPNOFQMRALPSCAL--CGGFTQSSGTLTSPGPPFY 703
 926 -----SADDLACGEILITESTGTLSPGHPNV 952
 704 PNNLCTWIIETSHGKGVFTFTPHLE-----SGHYLLITENGSTFQRLRLTSSRLPA 759
 953 PHQINCTWIIIVQPNHILHMETFTLBEHVNCTNDYIEVDYDSETS--LGYYCCSKSIP 1011
 760 PISAGLYGFTAQVRFISDPSMSYEGFNITFSEYDLPECEPEVAVYSIRKGLQGVGDT 819

1012 SLTSS--GN-SLMLVFVTDSDLAEGFLINYEAI----- 1042
 820 LTFSCFPGLYRFGTARITCIGRRRLMSSPLPRCAVACGNSVYTGCTLTSPFPVYNN 879
 1043 -----SAAITACI-----QDYTDIGTTSNPFNNYNN 1070
 880 NHECIYSIQTPGKGIQLARAFELSEG-----DVLKYVDGNNNARLLGVFSSHEMMG 933
 1071 NMECIYRITVTRTQQLIAVHFTNPSLEAIGNVYTDPLEIRDGGYKSPILGIF--YQSNLP 1129
 934 VTLNSSSLIWLDFITDAMNTSKGPELHSSPELLKCEDPGRPKGYKVADBGHFAGSSV 993
 1130 PTLTSSNKLMLKFSDDQIDTSSGFSAYVD-----GSSYGCQGNLTSSG 1174
 994 SFSQDPGY-----SLRGSPELL-----CLSGRRITMDRP--- 1022
 1175 TF-ISPNYMPYHSECECWMLKSHGSAFELFPDFLHEHNPCTLDVLAAYDSSNS 1233
 1023 -----LPTVAECGGTV-----RG-----EVSQVLS 1044
 1234 HLLTQLCGDEKRPPLIRSSGDSMFILKRTDEGQGGFGFAEYRQTCENVYIVNQTYGLES 1293
 1045 PGYAPYENHNCIWTIEAECTTIGLHFLVDEEVH-----DVLRIWDGVEGCVLLK 1099
 1294 IGYPNYSNQCNTTIRATTNTVNTYFLAFDLEH--HINSTDYLELDGPRQNG--- 1348
 1100 ELSGPALPKDLHSTNSVVLQFSTDFTSKQ--GFAIQFSV-----STAT--SCNDPGIP 1150
 1349 RYCGVDLPFGSTSSKQLVLLLTGQVGRREGFMQFVYVGGCELSGATGSFSPFP 1408
 1151 QNGSRSGDS--WEAGSTYFQCDPGYALQ-----GSAETSCVKI 1187
 1409 NRPYKKECIW-----YIRDPGSSIQTLTHDPVEYHRCNFDVLEIYGGPPHSPRI 1462
 1188 -----EN-----RF-----FQPSPPPTCIAPCGGLTSPS 1212
 1463 AQLCTGRSPENDMQVSSNGNELAIRFKTDSLINGRGNASQ-----AVTGCCGGIQAOPS 1518
 1213 GYILSPNTPPEPPECKEDCMKVTYVSPDYVALVFNIFENLEPGVFLHLYDGRDS--LSPLI 1271
 1519 GSHSPNYPSPYRSNTDCSWAIRVDRYHRLVLTDFDLPEQDCIMAYDGLSSTMSRLA 1578
 1272 GSFYSGQLPGRITSSNSGLFLAFPSDASVSAGVIVDTEPRRSC-----FD----- 1319
 1579 RTCGEQLANIVSSGNSLFRFOSGSPRQNRGRAOF---RACCGHILITSSFDVYSS 1634
 1320 ---PESINQTR---VGSIDLKSSVYTYCHGVEYEGSTLSC---IL--GPDGKPV 1366
 1635 PRFPANYPNNQNCMIIOAOPPLNHITLSFTH--FELERSTTCARDVLEILDGHEAPL 1692
 1367 -----MNNRPVCT-----APCGQYVGSQVVLSP 1392
 1693 RGRYCGTDMPHRTTSSSALTLPFVSDSSISAGGFHTTVAASACGTFPMAGIINSP 1752
 1393 NYPONTYSGQICLY--FTVPEADYVVFQGFAPFHTAL-----NDVVEVDSHQSRL 1444
 1753 GYDPITVPNNVCEWNIISSGNRL---QLSFISQLEDSQDCSRDPVEIRGNA-----T 1804
 1445 SSLSGSHGSLPLATNOV---LIFSAKGLAPARGFHYVQAV----- 1486
 1805 GHLVGRYCGNSFPLNYSIVGHTLMVAFIDGSGSGTGFQATPMKIFGNDNI VGHQVA 1864
 1487 -----PRTSAQO-----CSSVPEPRY----- 1502
 1865 SPFWPENYPPNNSYNQWVNVNASHVHGRILENDIEIONCYVDKRIYDGPSTHARLIG 1924
 1503 ---GKRLGDSFYGAIVRECNAGVALQGSRETECLVPBALAQW--NVSAPTCVP--- 1554
 1925 AYCGTQTESFSSGNSITLFFHYSDSSISGK-----GFLEWFAVADAPDVLTTIAP 1975
 1555 ---CGGNLTBRGT---ILSPGPPEPYLNSLNCVMKIVPEAGAGIIOVVSFTEQN--- 1605
 1976 GACGGFL--RTGAPVFLFSPGWDYSNRVDCVTWLDIAADST--VELNITLSDIESHRTC 2032

QY 1606 -MDSLVEVDGADNTVTMLGSPSGTTPALNLTSTNOCLHFFYSDI SVSAGFHELYKTVG 1664
 DB 2033 AYDYLVRDGDNNLAAQULAVLCGRBIPRISTGEBYFIRFTSSSVTRAFASFSH-- 2089
 QY 1665 LSSCPREPAPNSGVKTERYLVDNVSPQC-----EPGYALQGH-----AHISCM 1709
 DB 2090 -KSGCGYLHADRGLITTSKPYETPSNLNCSMHVLVQGLTIAVHFEQPPQIPNGDSSCN 2148
 QY 1710 PG---TYRRMN--YPPPL-----CIAO----- 1726
 DB 2149 QGDYLVRNGPDIYSPLPGPGNGHFGCSHASSTLFTSDNQMFVQFISDHSNEGQFKI 2208
 QY 1727 -----CGSTV-----EEMEGVILSPGPGNYPNSNDCSKIALPVGFGHIF--LN 1771
 DB 2209 KYEAKSLACGGNNTYIHDASAGYVTSRNHPNYPHADCIWILAPETRIQLQFEDRFD 2268
 QY 1772 FSTEPRN--HDYIERNGPYETSRMMGRPSGSELSLSTSHETTVYFHSQNRPGFK 1829
 DB 2269 IEVTPNCSNLYLELDGVDSDAPILSKFCGTSLPSSQMSGSEWVYLRFRSDNSFTYHFK 2328
 QY 1830 LEVAYAYELQECPPDEBPFGANGIVRAGY-----NGQSVPFECLEPGYQLT----- 1873
 DB 2329 AK---YSIAQCGRVPQSGGVESIGHPTLPYRDNLFCGEMHLOGLSGHYLTISFEDFNLO 2385
 QY 1874 ---GHVPLTQGHGNNRND-----HPLP----- 1893
 DB 2386 NSSG---CEKDFVEIWDNHTNSGILGRYCGNTIPSDITSNTAVRFTVDSVTASG 2440
 QY 1894 -----KCEVPCGNGITSSNGTVYSPGPPSYSSODCVMLITVPIGHVRL--NLSTLQ 1945
 DB 2441 FRLAFESMECCGDLQSGISTFTSPNPNPNPHGRICEMWITAPBGRRIITLPMNNRLA 2500
 QY 1946 TEPS--GDFTIWDGPGQTAPRLGVFTKSM--AKKTVSSNOVLKFKHDAAT--GGIFA 2000
 DB 2501 THPFCNNEHVIVFNGIRNSBPQLEKLCSSVAVNSNEIKSSGNTMKVIFFTDSDRBYG--FT 2559
 QY 2001 IAFAPAYELTCCPPTIILPNAVVTENEFNIGDIYRRCLEPGFLYNG----- 2048
 DB 2560 ASYTSSEDAVCGGS--LPN---TPEGNFT-----SPGYGVNRYSHNLNCEMTLS 2604
 QY 2049 -----EILTCKLGTVLOFEGP-----PRICE 2069
 DB 2605 NPNQNGSSISIHFEFDYFLESHQDFVLEBRVG--DADGFLMRLCGSPKPTLPVIR 2661
 QY 2070 -----VHCPTNELI-----TD-----STGVILSQSYPGSYPOFQTCSTWLRV 2106
 DB 2662 YSQVWIHFVTERVEHIGFAKYSFTDCGIGIIGDSGVITSBNYPNAVDSLTHCSWLEA 2721
 QY 2107 EPDYNISLTVYFLESKO---YDFELFDGPGSQSLKALSGNYAPLIVTSSNSVY 2162
 DB 2722 POGHTITLTFSDPFIETHTTACAMDSVYTRNGSPESPITIGQYCN--SNPRTIQSGSQVLV 2780
 QY 2163 LRWSSDAVYNRKGFIRYSAPYCSLPAPLHGFLIGOSTOPGSIHFGCAGARLVGHS 2222
 DB 2781 VTFPSDHSLOGGF-----YATWNTQT-----LQCGG---IFHS 2811
 QY 2223 MALCTRRPGGYHLMSEALPLCOALS--CGLPEAPRN--GMVFEKXYTVGKAVYSCSEBYHL 2280
 DB 2812 DNGTIRSBH---WPQNFENSRCSMTAITHKSXHLAISFNNNFLIPSDGQCCQNSFVKV 2867
 QY 2281 QAGGEATATAC-LDTGLMSNRNVPPOCVPTPDVSSISVEHGRWLLIFETQYQAOQLML 2339
 DB 2868 WAGTEEVDAKLLATGC---GNVAVG--FVITPSNTFAV-----POSGBA- 2907
 QY 2340 ICDGYVYTGQVRVRCQANGKMSLG-----BSTPTCRIISGCELPYPNGHRIG 2388
 DB 2308 ---PAQGSFASVFRCCSNFTGSPGYIISPNYKQYDNNMCTVY-----LEAN----- 2953
 QY 2389 TLSVYGAITAI-FSCNCGYTLVGSVRECMANGLWSGSEVRCLAGH-----CG--T 2435
 DB 2954 PLSTVLLTFVGFHLEARSATVGS---CVNDG-----VHIIKGYSVWSTPFATVCGDEM 3003

QY 2436 PEP1-VNQHNGENYSYRGSVVYOCNAGFRILIGMSVRIICQODHHMSGKTPCPVITCGHP 2494
 DB 3004 PAPITINGPVLNNTYSBQITDF---GFK---FSYH-----ISCGV 3040
 QY 2495 GNPVNGLTQGNQFNLDNVKFCVCPGY-MAEGARSOCLASGQMSDMLPTCRIINCTDPG 2553
 DB 3041 FNFSSGI-----ITSPAYSADVDPMDHCLVITITVS----- 3072
 QY 2554 HOENSVQVNASGPHRSFGTTVYSRCNHGFYLLGTIVLSCQGDGTMDRRPQCLLVSCG 2613
 DB 3073 ---DKVIELKRS---DFDVVPSTSCSHDYLA-----YDGNATSD---PLGKFCG 3114
 QY 2614 HPGSPHPSQMSG-----DSYTVGAVVRSYSGIGRTLVGNSTRMCGLDGHWGSLPH 2664
 DB 3115 SK-RPPVVKSSNNMSMLVFKTDSFTQTKAKKMSF--RQTL--GPQCCG--GYLTG--- 3163
 QY 2665 CSGTSVGVCDPGIPAHGIRLGDSPDPC-----TYMRPSCENGHYLR- 2706
 DB 3164 ---SNNTFASPDSDNGM-----YDKNLNCWIIIAVPVKV/IHLTENTFALBAASTROR 3214
 QY 2707 -----GSSERTQANGSWSGQPECGVJSCGNPGTSPNARVFSQGLVFSSSIVY 2756
 DB 3215 CLYDYVVLXDGDSENAMLA--GTFCGSTVPAPFISSGNFLT-----VQFISDLTL 3262
 QY 2757 ECREGYAT--GLSRHC--SVNGTWTG---SDPECLVINGCDPGIPANGRLIGNDFRYNK 2810
 DB 3263 E-REGFNATYITIMDPCGTYNATWTPONISSP-----NSSDPVPS----- 3304
 QY 2811 TVTYQCVPGYMESHRSVLSCTKDRTWNGTKPKYCKALMCKPPLIPNGKVGSDFMWGS 2870
 DB 3305 -----ICTWVIDISPRH--QQVKIT-----VMAL 3325
 QY 2871 SVT-YACLEGYOLSLPAVFTCEGNS---WTEBLPOCFPVFGD-----PGVPSR 2916
 DB 3326 QLTSDQCTQVY---LQLODPPQGHGNSRFQFCGNASAVVAFYSSMSTAMVIFKSGYVNR 3382
 QY 2917 GRREDRGSYR--SSVSFSCPLVLVGSPPRFQCSQDQTSCTDPLTLTTCADGVP 2975
 DB 3383 NSR--MSFTYQIADCNNDYHKAQNLKSP---GMPDNYDNDKQCTVTLTA---P 3428
 QY 2976 Q-----FGIONNSQGYOVGSYVLFRCQKGYL-----LOGSTFTCLPN- 3013
 DB 3429 QNHTISLPHSLIGIENSVE-----CRNDFLEVRNGSNSNSPLKCYCGTLLPNP 3477
 QY 3014 -----LWTSCTPPDCVPHHCQBPETHYANVAGALDLPBGY 3049
 DB 3478 VFSQNNELYLRFKSDSVTSRGEYELIWTSSPSC-----GCTLYGDRGSGFTSP--GY 3527
 QY 3050 -----TLTPARR-----ASPS----- 3061
 DB 3528 PGTYPNNTYCEWLVAVAGLVTINFYFISIDDPGDCVQNYVLLYDGPNASSPSGPGYCG 3587
 QY 3062 ---RVAPSTAPAR---RMAAGQASRPSAMR 3085
 DB 3588 GDTSIAPFVASSNQVFIKFIADYARRSAPR 3618

RESULT 3
 T08618
 Intrinsic factor-B12 receptor CUBILIN precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 11-Jun-1999 #sequence: revision 11-Jun-1999 #text: change 09-Jul-2004
 C:Accession: T08618
 R:Moestrup, S.K.; Koziyaki, R.; Kristiansen, M.; Kaye, J.H.; Rasmussen, H.H.; Brault, J. Biol. Chem. 273, 5235-5242, 1998
 A:Title: The intrinsic factor-vitamin B12 receptor and target of teratogenic antibodies
 A:Reference number: Z16459; MUID:98148073; PMID:9478979
 A:Accession: T08618
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3623 <MOE>
 A:Cross-references: UNIPROT:O70244; EMBL:AF022247; NID:G3834379; PIDN:AACT1661.1; PID:G C:Genetics:

A:Gene: CUBILIN
 C:Superfamily: Intrinsic factor-B12 receptor cubilin; EGF homology
 C:Keywords: egg yolk; endocytosis; glycoprotein; intestine; kidney; peripheral membrane
 F:1-20/Domains: signal sequence #status predicted <SIG>
 F:21-363/Product: intrinsic factor-B12 receptor CUBILIN #status predicted <MAT>
 F:133-164/Domains: EGF homology <EGF1>
 F:436-467/Domains: EGF homology <EGF>

Query Match 8.3%; Score 1404.5; DB 2; Length 3623;
 Best Local Similarity 20.6%; Pred. No. 9.2e-71;
 Matches 695; Conservative 378; Mismatches 1083; Indels 1219; Gaps 128;

```

QY 101 NEGVDLOGSKRITCKVKSMDFAVSDHAPVCRAMCAHLEGPSSGITT--SPNPPIQYDN 158
DB 443 NGGTICIDINGFTC-----DCTSSWTGYCQTPQAACGGLISGQTGFAMHSPN--DYIH 496
QY 159 NAHCWIIITANPSKVKITLAEFFDLE-----RGYDTLYGDCGQGDQKTVLYMSQNA 212
DB 497 NVNCFWIVRT--DEEKVLHVTFTEFDLSASNCPREY--LQIHDDSSADPPLGRY----- 548
QY 213 CSDSPHTPGSRIPESMGDIWRQKWTLEICRDISSDARSGSVKSPKTSNAVELVAPG 272
DB 549 C-----GSRPPGIIH-----SSANALYFHLYS 570
QY 273 TEIQSGCGDPGIPAYGRREGSRPHHGDYLFECOPAFELVQKATCQKNNOWSAKPG 332
DB 571 EYISGR--GFTA-----RWEAKLFE 589
QY 333 CVFSCFNFETSPSGVILSPNYPEDYGNHLHCWVILARPESRTHLPNDIVP-----QF 388
DB 590 CGGLITDNY-----GSTITSPGYPPGRCDCWQVLVNPNSLITFTGTSLSHSDCK 645
QY 389 DFLVIXDQATAEALVGTFSGNQLPSSITSGHVARLEFQDTHSTGRGFNITFTT----- 444
DB 646 DYLEIRDGPFPQDVLGKFTCSLSTPLKTTGPAPARHHSDBSTSGKHGIYTLTQSD 705
QY 445 -----FNNH-----ECDDPGVNPNGKRGDSLQSGSISF 474
DB 706 LDCGNATYTDGELLPLPSGPFHSRQCVYLITQAGOEIVINFTHVELSEOMGCSHTY 765
QY 475 L-----CDEGLTGOSETITCVLKEGSAVW-----NSAVLRC--EAPCGHLTSPS 519
DB 766 IYVODHSLARKICGNETLFPKRSVSKWIRLIDALVOKASFRADYQVACGGMKRG-E 824
QY 520 GTLISPGWPGFYKDALSCAWITAEOPGYPIKITFDREK-----TEVNYDTEVDRGTYS 575
DB 825 GFRSPFPYPAVPERATCRWTISQPRQVVLNFTDPIQISASCDPTDYEIIPSSVLGS 884
QY 576 PLIGVHGTVQVQFLSTSNVYLLPSTDSHSDIGQLRYETITLQSDHCLDPGILPVG 635
DB 885 PGNKPFSSNIPSTITSVNYLLVYTFVKSSSMENRGFTAKFSSDKLE----- 931
QY 636 QRHGNDFYVVALTFSCDSGYTLSDGEPLCEBNFQSRALPSCALCGGFIQSSGTL 695
DB 932 -----CGEVLTAFTGIIIE 944
QY 696 SPGEPPDFYNNINACTWIETSHKGVFTFTFHLB-----SGHDYLLITENGSPFTPLRQ 751
DB 945 SPGHPPVYPRGVNCTMHWVVOQRQLIRLEFSSFYLEFHVNYCTDYLEIVDTAQT-F-LGR 1003
QY 752 LITGSRIPAPISAGLYGFTAQVAFIDFSMSYEGFNITSEYVLECEBEVPAYSIKRG 811
DB 1004 YCKSKSLPSSLTS--NSNSTKLIFVSDSALAHGFIYNEAID-----ASSV--- 1047
QY 812 LQGVGVDTLTFSCEPGYRLEGATRTCLGRRRLMSSPLPRVAAECGNSVTGOGTILSP 871
DB 1048 -----CLVYTTDNFGMLSSP 1062
QY 872 NFPVYNNNNHCTYISQTOPGKGIQLKAPFELSE-----GDVLKVYDGNNSARLLGV 925
DB 1063 NFPNNNPSSWECEYRTTVGLNQIALHFTDFTLEDYFGSCVDPAFVIRDGYETSPLVGI 1122
QY 926 FSHSEMVGVTJLNTSSLMWLDFTDAENSTSKGPELHFSFELLKCEDPCTPRFGYKVHDE 985

```

```

DB 1123 YCGS-VLPPIITSHSKMLKFKSDALAKGFSAYW----- 1159
QY 986 GHFAGSSVSFSCDPGYSLRGSEELLCLSGERRTWDRPLPTVAECGGTVRGEVSGQVLS 1045
DB 1160 -----GSSGT-----CGGNLT--FTGTVLTS 1179
QY 1046 GYAPAYENHNLNCITWTEARAGCTIGLFLVPTDEEH-----DVLRIINGPVSQVLKE 1100
DB 1180 NYMPFYHSSECEYWRLEASHGSPFELEFQDFILEH--HPSCSLDYLAVFDPPTNSRLIDK 1238
QY 1101 LSGPALPKLHSTFNSVNLQFSTDFTSKQGFALQF-----SVSTATCNDNG 1148
DB 1239 LCGDTPAPIRSKMDVLLKATRDAGQQRGEINFRQGCNDVNVYNTKSGILESINYRN 1298
QY 1149 IPONGSRSDSWEA--GDSTVFQCDPGVALQSGAEISCVKIE-----NREFWQSPPT 1199
DB 1299 PYDKNQRCWMTIQATTGNTVNTYFLGFVDESVMNCSTDYVELYDGPQWNGRYCGNNMP 1358
QY 1200 CIAP-----CGGDLTGPSSVILSPNYPPEYPRCKED 1231
DB 1359 GATTGSQLHLVHTHDGINSGEKPKQWFTHCQGGEMSGTAGSFSPGPNYPHNKCT 1418
QY 1232 WKVTVSPDYVALVFNIFNLE-----PGYDFLHIYDGRDSLSPILGSFYG---SOLPRRIE 1284
DB 1419 WNIKVAPGSSIQLTTHDPDVEYHTSCNDLSLEIYAGLDPNSPRIADCSQSPANPMQVS 1478
QY 1285 SSSNSLFLAFRSDASVNAAGFVIDYTENPRESCEPFGSIKNGTRVSGDLKLGSSVYYCH 1344
DB 1479 STGNELAIRFKDSTLTNGFGNASWRAVP----- 1507
QY 1345 GGYEVEGTTSLCIIIGDPDCKPVMNPRPYCTAPCGGQYVGSDDVILSPNYPONTYSGQIC 1404
DB 1508 GG-----CGGIIQLSRGEIHSPPYNNYANMTEC 1536
QY 1405 LYFTVPKDYVFGQFAFPH--TALNDVVEVDHGSQHSRLLSLGSHTGESLPLATSNQ 1463
DB 1537 SWIIVERHRLVILNITFDLEAPDSCLRLMDGSSSTNARVASVCGRQGPSSIIASGNS 1596
QY 1464 VLKFSAGLAPARBFHYVQAVPRTSATQCSVBEPRYKRLSGDSFVGALVIRECNSG 1523
DB 1597 LFRFRSGSSSQNRFRAPFRE----- 1618
QY 1524 YALQSPETIECLPVGALAQWNVSAPTCVPCQGN-LTERGTILSPGPEPYLNSCV 1582
DB 1619 -----ECGGRIMDSSDTIFSPLYPHNYLHNQCS 1648
QY 1583 WKI-VVPRGAGIQIOVSF-----VTEQWDSLVEFDGADNTVTMLGSFSGTTVPALNST 1637
DB 1649 WIIEAQPPFNHITLSTHFOLQNSSTDCRDEVEILIDGNDYDAPVQGRYCGFSLPPIISF 1708
QY 1638 SNOLYHFYDPSVSAAGFHELYKTVGLGSSCEBPRAVPSNGVKTGRVLYNDVVSQCEPG 1697
DB 1709 GNALTVRVYDTSINSGFRAIYSA----- 1733
QY 1698 YALQGHAIHSCMPGTVRRMNYPPLCIAQCGGTVEMEGVILSPGPNYSPNMDCSWKI 1757
DB 1734 -----STSSCGGSPYTLIDGINSIDYPRADHPMAECVNI 1768
QY 1758 ALPVGFGHIOPLNFSTP-----NHDYIEINNGPEYTRMMGRPSGSELPSSLLST-SHE 1812
DB 1769 ASSPENRLQPLSLFNLNLSNCKNDKDFEILREG-NATGHILGRYCGNSLPQNVSAEGHS 1827
QY 1813 TTVFPHSDHSONRPGFKLEYO-----AY 1835
DB 1828 LMRVFSVDSGSGTGFGQARFKNITFGNNNIVCTHCKIASPFWPGKYPYNSNYKVVNVDAY 1887
QY 1836 -----ELQECDDPEPFANGIVRGAGYN-----VG-----QSYTFECLP 1868
DB 1888 HIHGRILMEMDIEPTNCFVSLKIYDGFDRSHRLIGYGCOTQESFSSNSLSLTFQSS 1947
QY 1869 GYOLTGHPVLTQCGITRNMDHPLPKCEVPCGNGITSNGTV--YSRQFSPYSSQDCV 1926

```

```

Db 1948 DSSVSGRGLLEWFAVDSDSTPTIAPGACGGFMWTDGPFVHIFSPGMPREYANGDCI 2007
QY 1927 WLTTPVIGHGVRLNLSLQTFP-----SGDFTTIDGSGQQAAPRLGVTRSAKTQVSSS 1982
Db 2008 WIIYAP-DSTVELNLSDLEPQOSCNVDKLIKXDGSDSLPDLAVICGVSPRPIRSTG 2066
QY 1983 NOVLKPHRDAATGCIPIALAFSAVPLTKCP-----PPTLPNAAVTVENE 2027
Db 2067 EYMIRTSIDTSVAG---TGFNASFHKSCGGYIHLADGVTSPKPTDLYLNL-----NC 2118
QY 2028 EFNIGDIVRYRCLPGLF-----VGNELTKLQTYLQF-EGP-----PPI----- 2067
Db 2119 SMHV--LVQ-----TGLTIAVHFEQPFQIQRNDSFCGQDYLVLNKGDNHSPPLGPGRN 2172
QY 2068 ---CEVHCPT-----NEL-----TDSTGVI 2085
Db 2173 GRFGGMWAPSTLFTSGNEMFVQFISDSNGQGFIRYKAKSLACGGTVYIHDSDSGYL 2232
QY 2086 LSGYPSGVPOFOTCSMLVRYEPDYNISLTVYFYLEKQYDEF-----ELFD 2132
Db 2233 TSPYTPANYPOHACITWILEAPRGRSIOLOFE-----DQFNIEDTPNCVSYLEIRD 2284
QY 2133 GPSGQSPBLKALSGNYSAPLIVTSSNSVYLRMSDHAYNRKGRKIRYSAPYCSLPAPL 2192
Db 2285 GANMANLVSRLCG-HTLPHSWSRSRRIYLFKHTDGSSTWYKAKYASLASC----- 2336
QY 2193 HGFLGCTSTOPGSHFGCNAGRLVGHSMALCTRHPOGYHLMSEALPLCOALSGLPE 2252
Db 2237 ---GGTVSGDSGVIR---SIGVPTLPYANV-----FCQMPRIQLP- 2371
QY 2253 APKGVWFGKEVYTGTAUVVSCSEGYHLQACAEATAECLDTGLWMSNR-----NV 2301
Db 2372 ---GHYLTIS-----FEDFNLOSFGCTKQFV--IENHTSGRVLCRYGNS 2414
QY 2302 PPOCVPTCPDVSSI-----SVHGWRILIFETQYQFOQLMLICDPGYVYTGQRYIRC 2355
Db 2415 TPSSVD-TSSNVAVKFVTDGSAAGRLDFKSSRQ-----VCG----- 2453
QY 2356 QANGKMSLGD-STYTCRIISCGELPIP---PNGH-RI--GTLSTVYGATAIF----- 2399
Db 2454 ---GDLHGPT-----GTFTSNYPNPNPBARICEMTIVQGRRLVLTFTMLRL 2499
QY 2400 ---SCNSGYLVGSRVRCMANGMWSGSEVRCLAGHCGPREPIVNHINGENSYGVS 2455
Db 2500 STOPSCHSEHLIV-----FNQIRNS-----PLQ----- 2524
QY 2456 VYOCNAGFRLIGMSVRICOQ-----DHHWSGKTFPCVPITCGHGNPNGLTQGNQNLN 2510
Db 2525 ---KCSRNVNNEFKSSGNTMKVVPFDG--SRPYGFT----- 2559
QY 2511 DVYKVCNPGYMAGAARSCCLASGQMSDMLPTCRIINCTDPGHQENSVR-----QV 2562
Db 2560 ---ASTSTEDAVCGG---FLPSVSGNFPSPGY--NGIRDYANRLDCW 2601
QY 2563 HASGPHRFSGTTSYCNHGYLLGTPVLSQ-----GDGTWDRPRQCLLVS 2611
Db 2602 TLSNPNKENSISIT-----YFLELSIBSHQCTDVLAFRVADGDLIEKFCLSLA 2653
QY 2612 CGHGPSPPHSQ-----MSGD--SYTVGAVVAYS--CIGKRT--LVGNSTRMCGLDGHW 2658
Db 2654 PTAFLVLPYRQVWHFVNSERVEYT-GFYIEYFTDGGIRITGNGVYISSNPNVLSAW 2712
QY 2659 TGSILPHSGSIVGCGDPGIRPAHGIKRGDSFDPTWRFSCGACHVLRGSSERTQANGS 2718
Db 2713 T---HCSWMLKABEG-----HTITL-----TFSDFLLEA-----HPTCTDSV 2747
QY 2719 W--SGSQPEGVIV--GKNP-----GTPSNARVVFSDGLVSSSIYVECEGQVATGLS 2769
Db 2748 TWRIGDPSGSPVIRYRGQSVPRIOGSGNOLIV-----TFNNQOQDTR-GFYA----- 2796
QY 2770 RHCSVNGTWTGSDDECLVINCDPGIRPANGRLANDFRVYNTLVYQCVGYMASHRVSV 2829
Db 2797 ---TWTTN-----ALGCGGTFHANG-----TIKSPHMPQTFPENSICSW 2833

```

```

QY 2830 LSCYKD-----RTWNGTKPVCKALMC-----KPPPLIP 2857
Db 2834 TVITHESKWEISFDSNFRIPSSDSQCQNSFVKWBSRLMINKTLLATSCGDVAESPITV 2893
QY 2858 NGKVVSGDFPMWGSSTVYACLEGYQLSLPAVFCE--GNGSWTGEILPQCFPVFCGDPG-- 2912
Db 2894 SGNI-----FTAVFOGBEAMAGQFASFSIRCGRTFTNTPSGDI 2931
QY 2913 -VPSRGRREDRGS-----YRNSVSPSCHPLVLVGSPPRRFCQSDG----- 2952
Db 2932 ISPNFPKQYNNMNCYTLIDADPQSLVILTFVFSHLEDRSAITGT-----CDHGLHITIK 2987
QY 2953 ---TWSGNQ---PSCID-PTLTT--CABPGVRFQIQQNNSQGVGYSTVLFRCQK 2998
Db 2988 RNLSSTPLVITCSEETRLPTVDGVLNFPYSDAYTTDFGFKISTYRAITCGG--TYNSS 3045
QY 2999 GYLLOGSTTRTCLPN 3013
Db 3046 GILRSPSYSYNPN 3060

```

RESULT 4
136936
complement receptor 1 - chimpanzee (fragment)
C:Species: Pan troglodytes (chimpanzee)
C:Date: 02-Jul-1996 #sequence, revision 02-Jul-1996 #text, change 09-Jul-2004
C:Accession: I36936; I36937
J:Blottingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.
J:Immunol. 153, 691-700, 1994
A:Title: Primary sequence of an alternatively spliced form of CR1. Candidate for the 75
A:Reference number: I36935; MUID:94292799; PMID:8021505
A:Accession: I36936
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2014 <BIR1>
A:Cross-references: UNIPROT:Q29530; GB:I24920; NID:g551564; PID:AAA51438.1; PID:g55772
A:Cross-references: GB:I24921; NID:g557726; PID:AAA51439.1; PID:g557727
C:Genetics:
A:Gene: CR1
C:Superfamily: Cdb-binding protein alpha chain; complement factor H repeat homology
C:Keywords: duplication; glycoprotein; tandem repeat
F:18-74/Domain: complement factor H repeat homology <FH01>
F:79-136/Domain: complement factor H repeat homology <FH02>
F:141-207/Domain: complement factor H repeat homology <FH03>
F:213-268/Domain: complement factor H repeat homology <FH04>
F:272-328/Domain: complement factor H repeat homology <FH05>
F:333-391/Domain: complement factor H repeat homology <FH06>
F:396-462/Domain: complement factor H repeat homology <FH07>
F:470-526/Domain: complement factor H repeat homology <FH08>
F:529-586/Domain: complement factor H repeat homology <FH09>
F:722-778/Domain: complement factor H repeat homology <FH10>
F:1041-1107/Domain: complement factor H repeat homology <FH11>
F:1172-1228/Domain: complement factor H repeat homology <FH12>
F:1233-1291/Domain: complement factor H repeat homology <FH13>
F:1396-1362/Domain: complement factor H repeat homology <FH14>
F:1432-1489/Domain: complement factor H repeat homology <FH15>
F:1625-1681/Domain: complement factor H repeat homology <FH16>
F:1749-1815/Domain: complement factor H repeat homology <FH17>
F:1823-1879/Domain: complement factor H repeat homology <FH18>
F:1884-1940/Domain: complement factor H repeat homology <FH19>

Query Match 8.2%; Score 1390; DB 2; Length 2014;
Best Local Similarity 22.2%; Pred. No. 2.7e-70;
Matches 544; Conservative 266; Mismatches 820; Indels 822; Gaps 111;
798 CEEPEVPA---YSIKRGLQGVGDTLTFSCFPYRLGRTARIC-LGGRRLWSSPLPRC 853
Db CRNPDPVNMVHVIQIQF--GSQIKYSCTKYRILJGSSASACIIISGDIVINDNERPIC 136

```

QY 854 -VAECGVTGTGTLSPNFPVNNNNHCEIYSIQTOPGKIOLKARFELSGLVKY 912
Db 137 DRICGLPPTITNDFTSTN---RENHYGVSVYTRCNPGSG---GRVPEL-VGEP-SI 188
QY 913 YDGNNSARLLGVFSHSEMGMVLTNSTSSLMDFITDAENTSXGFEHFESEFELICED 972
Db 189 YCTSNDDQ--VGIMSGPAPQCIIPN-----KCTP 215
QY 973 PGTRFGYKXVDEBHF--AGSSVSFSCDPGYSILGSEBELCLSGERTWDRPLEFTCAEC 1030
Db 216 PNVEN-GILVSDNLSLSLNEVVEFRCPGFMKGPVRKQALNK--WPEBELFSCSRVC 272
QY 1031 GGTVRGEVSGQVLSPPYAPAYEHNLNCIWTIEAGCTTGLHFLVPTBEVHDPLRWGD 1090
Db 273 -----QPPP-----DVLH-----280
QY 1091 PVESGVLKELSGPALPKDLHSTNSVYLOFSTDFTSKOGFALQFSVSTATSCNDPGLP 1150
Db 281 -----A 281
QY 1151 QNGSRSGDSWEGADSTVFQCDPGYALOGSAEISCVKIENRFWQSPPTCIACGGDLTG 1210
Db 282 ERTORDXDNFSPGQEVFYSCEPGYDLRGAASLRCTPGD--WSPATPCEVXSCDDFMG 338
QY 1211 P--SGVILSPVYPPYPPGKECDWKVTSPDYVIALVFENLEPGYDFLHIYDGRDLS 1268
Db 339 QLNGRVLFP-----VNLQAKXDPVCD-----362
QY 1269 PLIGSPFGSQLPRIBESSNSLFLA---FRSDASVSNAGFYVDYTNPRESCFDPGSI 1323
Db 363 -----EGFQLKG---SSASCYVLAMESLMNSVPCBOIF-----CPSPPVY 402
QY 1324 KNGTRVSGDLKL---GSSVYTYCH---GGYVEBSTSLSCILGDPGKPVNNRRPVC 1374
Db 403 PNGRHKTPLLEVFPFGAVNYTCDPHDRGTTFDLIGESTIRCTSDQGGVMSSPAIRC 462
QY 1375 TAPCGQYVSGDVLSPNYPONTYSGQICLYETVPKDYVVEGFAFHTALNDVEVH 1434
Db 463 -----GIL-----465
QY 1435 DGHSQ--HSRLSSLSGSHTEGSLPLATSNOLIKESAKGLADARGF---HYVOAV 1486
Db 466 -GHQOAPDHPLEFKLTQTNASDFPIGTS---LTKXCREYVYGRFSTICTDLNLWMS- 519
QY 1487 PR--TSATQCSSVBERYK-RLGSPSVGAIIVRECNISGVALQSGPEIECLVPQALQ 1543
Db 520 PKOVCKRKSCKTPDPVNGMVHVTIDIOVSRLNYGCTGHRLLGHSSABCI-LSGNAH 578
QY 1544 WNVSAFTCV--VPCGMLTERRTILSPGPEPYLNSLNCWKIVBEGAGIOIOVVSFT 1602
Db 579 WSTKPPICORIPCGLPPTIANGDPISTNRENFHGS-----VVTYNC 620
QY 1603 EQWMSLEVEDGADNVTMLGFSFGTTVPALLNSTNOLYHMYSDISVSAAGFHLXYT 1662
Db 621 NPSRGRKATFE-----LVGEPSS-----LYCTND-----DQ 646
QY 1663 VGLSSCPER--AVPS-----NGYKTER--YLNDVVSFOCEPGYALOGHAHISC 1709
Db 647 VGIMSGPAPQCIIPNKTCPNVNENGLVVDNLSLFLNEVVEFRCPGFMKGPVRKQ 706
QY 1710 PGTVRWMNTPPPLCIQCGGYVEEMGVILSPGPNYPSNMDCSKIALPVGFGAHIQF 1769
Db 707 --ALNMEBELPSCSRVC-----722
QY 1770 LNFSTEPNHYEIRNGPYETSMRMRFGSELPSLSLSTSHETTYVYRHSDSQRPGK 1829
Db 723 -----QPPEDVL-----HAERTOR-----736
QY 1830 LEYQAVELQECDDPEBPANGIVRGAGVNGQSTFECLPGYOLTGHPVLTQCHGNRRMD 1889
Db 737 -----DKDNF-----SPGQEVFYSCEBGYDLRGAASLRCT--PQGDMS 772

```

```

QY 1890 HPLKCEV--PCGNITS--SNGVYSPGPPSPSSODCWMLITVPIGHGVRNLSTLQTE 1947
Db 773 PAAPCEVXSCDGMQOLNGRVLP-----VNLQAKV-----807
QY 1948 PSGDFTIWDGPQOTAPRLGVFTSMMAKKTVOSSNOVLLKTHRDAATGITALAFSAVP 2007
Db 808 ---DFVCD--EGFQLKG-----SSASCYV-----AGMSLWMSVP 839
QY 2008 LTK---CPEPTILPNAEVNTEEN--EENIGDIYRCLP-----GFTLVGNEILTKL-- 2055
Db 840 VCEQIFCPSPPVYIPNHRHTGKPLEVFPFGKAVNYTCDPHDRGTTFDLIGESTIRCTSDP 899
QY 2056 ---GTYLQFEGEPPICEV--HCPTEBLTDSIGVLSQSPGSIYQFQCSMLVREDEPY 2110
Db 900 QGNGW---SSPAPRGILGHCOA-----PD- 922
QY 2111 NISLVEYFLSEKQYDEFIFDPGSGQSPBLKALSGNYSAPLITSSNSVYLRMSDHA 2170
Db 923 -----HFLFAKLKTQTNASDFPIGTSLKYCEBRYGRPSTICLDNV--WSSPKD 972
QY 2171 YNRKGFKLRYAPYCSLPPAPLHGFILQSTQTOPGSIHFGNAGYRVLGSHMAICTRHP 2230
Db 973 VCR-----KSKCTPPDPVNGMVHVTIDIOVSRLNYGCTGHRLLGHSSABCIISG 1024
QY 2231 QGYHLMSEALPLCOALSGLPEAPKMGVFG--KEYVTGTAVYSCBGRH---LQAG 2283
Db 1025 NSAH--WSTKPPICORIPCGLPPTIANGDPISTNRENFHGSVYTRCNLSGRKRVFELV 1083
QY 2284 ABATAECL---DTGLMS-----NRNVP-----OCVP 2307
Db 1084 GEPISYCTSNDDQVIMSGPAPQCIIPKKTCPNVNENGLVSDNLSLSLNEVVEFRCP 1143
QY 2308 -----VTC-----PVSSIS-----VEHRMLIFETQYFOALMLICDP 2343
Db 1144 GFVMKGPVRKQALNKWPELFPSCSRVCOPPELHETHPSHODNPSPGQEVYVSCBP 1203
QY 2344 GYVYTGORVIRCOANGKWSLGDSTPTCRIISG---GELPIPPNHRIGTSLV--YGAHAI 2398
Db 1204 GYDLRGAASLHCTPGDMS--PEARCAVXSCDDFLQGL---PHGRVLPPLNLQAKVYS 1258
QY 2399 FSCNSGYTLVSGRYNECNANG---LMSGSEVRCLAGHGTPEPYNGHNGENYS---YR 2452
Db 1259 FVCEGDFRLKSSVSHCVLWGRSLMNSVPCBQIFCPNPPALINGHHTGTPREDIYFG 1318
QY 2453 GSVVYQCN-----AGFRLIGMSYRICQDDH---WSGKTPFC-VPTTCGH-----2493
Db 1319 KEISTICDHPHDPKMTFNLIGESTIRCTSDPHGNGVMSPPAPRCLEPVHAGHCKTPEQFP 1378
QY 2494 ---PGNPVNGLTQGNQFNLNDVRFVCNPGYMAEGAARSQCLAGQMSDMLPTCRIINCT 2550
Db 1379 FASPTIPIINDP-----EPVYGSILMYECPRGYFGKMFSTIS-CLENLWMSVSDNCRKSCG 1433
QY 2551 DPGHDENSVROVHASGPHRSFGTTVSYRCHGFYLLGTPLVSCQGDG---TMDRPPDQ 2607
Db 1434 PPEPFGNOMAHIND---IQFGSTVNYSCNEGRLLIGSPETTLVAGNNVTMDKAPIC 1489
QY 2608 LTVSGHGHSPPHMSGDSYT-----VGAVVRYS-----IGKTLVGN 2648
Db 1490 EITIC---EPPTISNGDFLNNRASFHNGTAVTYOCHTGPDGEOLFEVIGERSIY--C 1543
QY 2649 TRMCGLDHMTGSLPHCSGTSVGVCGDPIPAHIGIRLGDG---FDPGVNRFSCBAGHVL 2705
Db 1544 TSXDQYGVMSPPPRCISTN--KCTAPV--ENAIRIPGRKSRFSLIEIYRFRQPGFVM 1600
QY 2706 RGSSEPTCOANGSVSGQPECGVISCNP-----GTPSNARVVFSDGLVFSSSIYEC 2758
Db 1601 VGSHTVOCTNGRGPRLPHCSRV--COPPEILHGEHTPSH-----QDNFSPGQEVYVSC 1654
QY 2759 REGYVATGLSRHCSVNGTWGTGSDPECLVINCDD--PGIANGRLGNDPRYKNTVYQC 2816
Db 1555 EPGYDLRGAASLHCTPGDMSPEAPRCTVKSDDFLQGLPHGRVLPPLNLQAKVYSFVC 1714
QY 2817 VPGYMESHRVYS--VLSCTDRTWNGTKPVCKALMCKPPLIPNGKVYGS---DFMGGSS 2871

```



```

Db      1715 DEGRFLKRSASHCVLAQMK-ALMNSVPCVEQICRPAIILNHRHGTFRGDIPIKYE 1773
      2872 VTYAC-----LEGYQLSLPAVTCBGNWSWTEBLPQC--PFFVCGDGVPRRG 2917
      1774 ISYACDTHPRDGMFTNLIG-SSSIRCTSDPQNGWNSPAPRCBLVPAACPHP--PKIQ 1830
      2918 RREBRG-----FSYRVSFSFCHPPLVVGSPRRFCQSDGTWMSGQPSCIDPTLTTCADP 2972
      1831 NGHIDIGVSLYLPQMTTSTYICDPEGLVVGKGFICTTQGLWSQLDHYCKE---VNCSPF 1887
      2973 GVPOFGIGNN---SQGYVGSSTVLFRCQGYLLQGSTTRTCLPMLTWSGTFP 3021
      1888 -LPMNGISKELEMKVHYGVYTLKCEDGYTLBESSPWSQCADRMD--P 1936

RESULT 5
A43526
Complement C3d/Epstein-Barr virus receptor 2 precursor - mouse
N:Alternate names: complement receptor type 2
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A43526; A43538; A43215; A45802; B32215
R:Flngerolth, J.D.
J:Immunol. 144, 3458-3467, 1990
A:Title: Comparative structure and evolution of murine CR2. The homolog of the human C3d
A:Reference number: A43526; MUID:90229735; PMID:2139457
A:Accession: A43526
A:Molecule type: mRNA
A:Residues: 1-1025 <F1N>
A:Cross-references: UNIPROT:P19070; GB:M35684; EMBL:J04153; NID:G192687; PID:AAA37448.1
R:Flngerolth, J.D.; Benedict, M.A.; Levy, D.N.; Strominger, J.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 242-246, 1989
A:Title: Identification of murine complement receptor type 2.
A:Reference number: A43215; MUID:89098890; PMID:2783485
A:Accession: A43215
A:Molecule type: mRNA
A:Residues: 343-401/991-1025 <F12>
A:Cross-references: GB:J04153
R:Kurtz, C.B.; Paul, M.S.; Aegerter, M.; Weis, J.J.; Weis, J.H.
J. Immunol. 143, 2058-2067, 1989
A:Title: Murine complement receptor gene family. Identification and characterization of
A:Reference number: A45802; MUID:89381350; PMID:2528587
A:Accession: A45802
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'E', 100-101, 292-961, 964-1025 <KUR>
A:Cross-references: GB:M9281; NID:G192685; PID:AAA37447.1; PID:G387131
A>Note: the authors failed to translate GGA for residue 421 as Gly, and CCA for residue
A>Note: the authors translated the codon CAC for residue 727 as Asn
C:Superfamily: Complement C3d/Epstein-Barr virus receptor; complement factor H repeat hc
C:Keywords: alternative splicing; duplication; glycoprotein; receptor; transmembrane pro
F:1-11/Domain: signal sequence #status predicted <SIG>
F:11-105/Product: complement C3d/epstein-Barr virus receptor 2 (15-repeat form) #status
F:12-973/Domain: extracellular #status predicted <EXT>
F:14-73/Domain: complement factor H repeat predicted <FHO1>
F:82-138/Domain: complement factor H repeat homology <FHO2>
F:146-202/Domain: complement factor H repeat homology <FHO3>
F:207-263/Domain: complement factor H repeat homology <FHO4>
F:266-334/Domain: complement factor H repeat homology <FHO5>
F:343-358/Domain: complement factor H repeat homology <FHO6>
F:402-458/Domain: complement factor H repeat homology <FHO7>
F:463-514/Domain: complement factor H repeat homology <FHO8>
F:519-585/Domain: complement factor H repeat homology <FHO9>
F:594-649/Domain: complement factor H repeat homology <FHO>

```

```

F:654-704/Domain: complement factor H repeat homology <FHI2>
F:709-769/Domain: complement factor H repeat homology <FHI3>
F:778-833/Domain: complement factor H repeat homology <FHI4>
F:841-897/Domain: complement factor H repeat homology <FHI5>
F:902-958/Domain: complement factor H repeat homology <FHI6>
F:968-989/Domain: transmembrane #status predicted <TM>
F:990-1025/Domain: intracellular #status predicted <INT>

Query Match      5.3%; Score 906.5; DB 1; Length 1025;
Best Local Similarity 27.4%; Pred. No. 2.6e-43;
Matches 255; Conservative 130; Mismatches 414; Indels 133; Gaps 36;

      2183 PYCSLPAPLHGLFLLGDTSTQPGSIHFGNAGYRLVGHSMALGTRHPDGHLSSEAPL 2242
      26 PYSLP-----IYGVTLVLYTSPSRLIGKAFICISGNQHATDKAPPI 72

      2243 COA-----LSCGLPAPKNGWVFGKE--YVGTAKAVSCSEGYHLQAGAEATFELDTGLW 2296
      73 CESVNTKITSQDPIVPGGFNMKSKAPFRHGDSTVTFCKANFTWK--GSKTVMQANEMW 130

      2297 SNRNV--PCVPTCTCDVSSISVEHGRMLIFETQIQFQALMLI--CDPGYTTGQRV 2352
      131 GPRLPYCESDPFLECSPLPIHNGHHTGQHV---DQFVAGLSVTYSCEPGYLLTGKXT 186

      2353 IRCQANKWLSGDSPTCRITISGELPIRPNGHIGTLV-YGATFAPSCNSGYTLVGR 2411
      187 IKCLSSGDWD--GYTPCKAQCSPKPGKFPBGVKEPLSLQVGTYYFSCNEGTLQSQOP 244

      2412 VRECM--ANGWGSSEVRCLAGHGTPEPIVNGHNG-----ENYSYRGSVYVQCN----- 2460
      245 SSQGVIVYQKAIWTKKV--KEILCPRPVPRNHSHGSENVPGSTVYTTDDBPEK 303

      2461 -AGRLIGMSYRVC---QODHMSGKTPGV---PITGHP--GNPYNGLQGNQF 2507
      304 GVSPTLLGKTKINTTGSOKTIGWSGAPVCLSTSAVLCLQPKIKRQLSLIK--DSY 361

      2508 NLDNVFVFCNPGMAAGARSOCLASGWSMDMPTC-----RIINCDPQHOENV 2559
      362 SYNTVAFSCPGTLLGNRSIRCNAGTBPPRPVCEKGCQAPKLIIN---GQKDSY 417

      2560 ROYVASGPRHFSFGTYSYRCHNGFYLLGTPIVLSQCGDGTDRRPPOLVSCGHPSP 2619
      418 LL-----NPDSTISRYGCDPEYLLVGBDTHCTPEGKMTPTTPQCTVAECKPVG--P 468

      2620 HSQMSGSIYVGAIVRYSCIGKRTLVGNSITRMCGLDGHTGSLPHCGSTVVCGPPIP 2679
      469 HLFRRPONQFIRTVANSSCEGFLSSEAVQLCGTIPMLEIRLCKEIT---CRPPVI 525

      2680 AHGIRLDSFDP---GVNMFPSG---EAGHVLRGSSERT-----QANGSWSGSGPEC 2726
      526 HNGHTWSSSDVPGVGVVYVMCPGPEBGKFLIGEDYTHCTSDSRGSGSWSPAPLC 585

      2727 GV-----ISGNPQTPSNARVVPDGLVF-SSSIYVECREGYATGLSRHCSVNGTGS 2781
      586 KLSLPANQCTDVHVENVKLTNDKAPRYFVNDVWFKDDDEYLLSSGQIQCKANNITDPE 645

      2782 DPECLVINCDP---GIPNGLRLGNDPRYKNTVTVQCVGYMESHRVSVLSTYDRT 2837
      646 KPLCKEKGK--EPMKHGLP-----DSDHIKLVKTKCQNGYQLTYTEKQANNAENG 696

      2838 WNGKRPVCKALMCKPRLINGKVV--SDPMGSSVTVACLEGYOL-----SLPAFTC 2890
      697 WPKKIECTVILCQPPKXINGGHTGMAGKFLVGNVSYECDEGFYLLAEKSLQCVNDS 756

      2891 EGNWSWTGELPQCFP---VFCGDPGVPS--RGRREBRGSGSYRVSFSCHPPLVLVGS 2944
      757 KGHSGWGPRLQGLQSSPLNHCDBPYKHGKYLKNTKHSASNDIYHVFVNCQGINVGS 816

      2945 RRFQSGDGTWSGTQPSCIDPTLTTCADPVGVRGQIGNNSGQ---YGVSTVLFRCQGY 3000
      817 LIRCHTNWLPVGVPTCIRASLQCSQSPSTIPNG--NHTGSIARFPFGMSVMVSCYGF 874

      3001 LLQSTTRTCLPMLTWSGTRPDCVPHNCRQPE 3032

```

Db 875 IMAGEARLICTHEGTSQPPFCKEVCNCSFPE 906

RESULT 6

complement C3d/Epstein-Barr virus receptor precursor - human

N/Alternate names: complement receptor 2; CR2/CD21

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1992 #sequence revision 07-Jul-1995 #text change 09-Jul-2004

C/Accession: J10028; A39958; A24319; B24319; C24319; D24319; E24319; F24319; PLC

R/Weis, J.J.; Tothaker, L.E.; Smith, U.A.; Weis, U.H.; Fearon, D.T.

J. Exp. Med. 167, 1047-1066, 1988

A/Title: Structure of the human B lymphocyte receptor for C3d and the Epstein-Barr virus

A/Reference number: J10028; PMID:88171282; PMID:2832506

A/Accession: J10028

A/Molecule type: mRNA

A/Residues: 1-1091 <WRI>

A/Cross-references: UNIPROT:P20023

A/Note: nucleotides 1566-1625 are missing from Figure 1; therefore, residues 522-542 have

R/Moore, M.D.; Cooper, N.R.; Tack, B.F.; Nemerow, G.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 9194-9198, 1987

A/Title: Molecular cloning of the CDNA encoding the Epstein-Barr virus/C3d receptor (com

A/Reference number: A39958; PMID:88097454; PMID:2827171

A/Accession: A39958

A/Molecule type: mRNA

A/Residues: 1-456, 'G', 457-644, 'R', 646-669, 'R', 671-816, 'NCSAEVILKAMILERA', 835-840, 'L', 84

A/Cross-references: GB:J03565; NID:G181919; PID:AA35784.1; PID:G181920

R/Fujisaku, A.; Hartley, J.B.; Frank, M.B.; Gunner, B.A.; Frazier, B.; Holers, V.M.

J. Biol. Chem. 264, 2118-2125, 1989

A/Title: Genomic organization and polymorphisms of the human C3d/Epstein-Barr virus rece

A/Reference number: A32036; PMID:89123277; PMID:2563370

A/Accession: A32036

A/Molecule type: mRNA

A/Residues: 1-456, 'G', 457-658, 'I', 659-718-1050, 'I', 1052-1060, 'E', 1062-1091 <FLU>

A/Cross-references: GB:J04463

R/Meis, J.J.; Fearon, D.T.; Kijckstein, L.B.; Wong, W.W.; Richards, S.A.; De Bruyn Kops,

Proc. Natl. Acad. Sci. U.S.A. 83, 5633-5643, 1986

A/Title: Identification of a partial CDNA clone for the C3d/Epstein-Barr virus receptor

of complement.

A/Reference number: A94114; PMID:86287311; PMID:3016712

A/Accession: A24319

A/Molecule type: Protein

A/Residues: 226-230, 'XIQ', 257-267, 332-341, 583-591, 'Q', 593, 'D', 595-596, 728-735 <WE2>

A/Experimental source: B-lymphoblastoid cell lines SB and Raji

C/Genetics:

A/Gene: GDB:CR2

A/Cross-references: GDB:119802; OMIM:120650

A/Map position: 1q32-1q32

C/Subfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat ho

C/Keywords: alternative splicing; duplication; glycoprotein; transmembrane protein

F/1-20/Domain: signal sequence #status predicted <Sig>

F/21-1091/Product: complement receptor 2 (16-repeat form) #status predicted <MMT1>

F/21-658/Domain: complement factor H repeat receptor 2 (15-repeat form) #status predicted <MAT

F/23-82/Domain: complement factor H repeat homology <FH02>

F/91-146/Domain: complement factor H repeat homology <FH02>

F/154-210/Domain: complement factor H repeat homology <FH02>

F/215-271/Domain: complement factor H repeat homology <FH03>

F/276-442/Domain: complement factor H repeat homology <FH04>

F/351-406/Domain: complement factor H repeat homology <FH05>

F/410-465/Domain: complement factor H repeat homology <FH06>

F/470-521/Domain: complement factor H repeat homology <FH07>

F/526-592/Domain: complement factor H repeat homology <FH08>

F/601-656/Domain: complement factor H repeat homology <FH09>

F/660-716/Domain: complement factor H repeat homology <FH10>

F/720-772/Domain: complement factor H repeat homology <FH11>

F/777-837/Domain: complement factor H repeat homology <FH12>

F/846-901/Domain: complement factor H repeat homology <FH13>

F/909-965/Domain: complement factor H repeat homology <FH14>

F/970-1026/Domain: complement factor H repeat homology <FH15>

F/1034-1056/Domain: transmembrane #status predicted <TM>

F/1057-1091/Domain: intracellular #status predicted <INT>

F/121,127,294,372,622,698,858,881,919/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 5.3%; Score 904; DB 1; Length 1091;

Best Local Similarity 25.2%; Pred. No. 3,9e-43;

Matches 281; Conservative 133; Mismatches 438; Indels 264; Gaps 49;

QY 2012 PPPTLPNNAVTVENEFNIGDIVRYRCLPGTLVNEILTC---KL-GTVLQFGSPP 2066

Db PPPTLPNNAVTVENEFNIGDIVRYRCLPGTLVNEILTC---KL-GTVLQFGSPP 80

QY 2067 IGVHCPNTELLTDSTGVILSGSYSPYQFQTCGMLVREBDYNISLTVEYFLSEKQYD 2126

Db IGVHCPNTELLTDSTGVILSGSYSPYQFQTCGMLVREBDYNISLTVEYFLSEKQYD 94

QY 2127 EPEIFPDGSGSPLKALSGNYSAPLIVTSSNSVYLRMSDHAHYRKQFKTRYAPYCS 2186

Db EPEIFPDGSGSPLKALSGNYSAPLIVTSSNSVYLRMSDHAHYRKQFKTRYAPYCS 108

QY 2187 LPRAPLHGFILGOTSTPGSGIHFQGNAGYRLVGHSMALCTRHPOGYHLSBA-IPLQCA 2245

Db LPRAPLHGFILGOTSTPGSGIHFQGNAGYRLVGHSMALCTRHPOGYHLSBA-IPLQCA 148

QY 2246 LSCGLP-EAPKNGMFGKEYT-----VGTAAVSCSEGYHQAQAEATACIDTGLM 2296

Db LSCGLP-EAPKNGMFGKEYT-----VGTAAVSCSEGYHQAQAEATACIDTGLM 203

QY 2297 SNRNPPOCVPTC-----PDVSSISVHEGRMLIFETQYQFOQMLICDP 2343

Db SNRNPPOCVPTC-----PDVSSISVHEGRMLIFETQYQFOQMLICDP 244

QY 2344 GYYTGGVYRIRCOANGKMSLGDSTPTCHIIISGELPIPP---NGHRIGLTV---YGATA 2397

Db GYYTGGVYRIRCOANGKMSLGDSTPTCHIIISGELPIPP---NGHRIGLTV---YGATA 301

QY 2398 IFSGNS-----GYTVSSRYARECAN---GLMSGSEVRC---LAHGCTPEPIYNGH 2443

Db IFSGNS-----GYTVSSRYARECAN---GLMSGSEVRC---LAHGCTPEPIYNGH 361

QY 302 ITTCDPDEEGVNFILIEESTLCTVDQKGTWSGPARCELSAIVQCHPPIILGRM 361

Db ITTCDPDEEGVNFILIEESTLCTVDQKGTWSGPARCELSAIVQCHPPIILGRM 2501

QY 2444 ING---ENYSYSGSVVYQCNAGFRILIGMSVRICQDQDHMSGKTPFCVPTTCGHPGPNVGL 2501

Db ING---ENYSYSGSVVYQCNAGFRILIGMSVRICQDQDHMSGKTPFCVPTTCGHPGPNVGL 420

QY 362 VSGQDRRTYNDYVIFACMFGLTKGSKOITCMAQGTPEBAPVCEK-EQCAPPIITLQGO 420

Db VSGQDRRTYNDYVIFACMFGLTKGSKOITCMAQGTPEBAPVCEK-EQCAPPIITLQGO 2558

QY 2502 TQGN---QFNLDVVKFVCPNPGVMAEGARSOCLASGQMSDMLPTCRITINCTDPGHQENS 2558

Db TQGN---QFNLDVVKFVCPNPGVMAEGARSOCLASGQMSDMLPTCRITINCTDPGHQENS 474

QY 421 KEDRMHVFDDGTSTIKYSNCNGVYLVBESIQ-C-TSEVWTPEVPQCKVAAAEANG----- 474

Db KEDRMHVFDDGTSTIKYSNCNGVYLVBESIQ-C-TSEVWTPEVPQCKVAAAEANG----- 2616

QY 2559 VROVNASGPHRPSRETTYVYRCNNGFYLLGTPVLSQCGDGTWDBRPQCLLVSCGHPG-- 2616

Db VROVNASGPHRPSRETTYVYRCNNGFYLLGTPVLSQCGDGTWDBRPQCLLVSCGHPG-- 532

QY 475 -RQLTLTPQHOP-VRPDVNSSCGEGYKLKSGSVYQECQGTIPMEIRLCKEITCPPEPVI 532

Db -RQLTLTPQHOP-VRPDVNSSCGEGYKLKSGSVYQECQGTIPMEIRLCKEITCPPEPVI 2665

QY 2617 -SPHSQMSGSDYTVGAVVRYSCIG-----KRTLVGNSTRMGCLD-----GHMTGSLPHC 2665

Db -SPHSQMSGSDYTVGAVVRYSCIG-----KRTLVGNSTRMGCLD-----GHMTGSLPHC 592

QY 533 YNGAHTGSLDLPFGYGTITVYTCNPGPRGVGFELIGESTIRCTNSDOERGTWGPAPLC 592

Db YNGAHTGSLDLPFGYGTITVYTCNPGPRGVGFELIGESTIRCTNSDOERGTWGPAPLC 2721

QY 2666 SGTSTGVV-CGDPGIPAHGIRLGDSPD--GTVMRFSCGAGVLRGSSERTCOANGNSG 2721

Db SGTSTGVV-CGDPGIPAHGIRLGDSPD--GTVMRFSCGAGVLRGSSERTCOANGNSG 651

QY 593 KLSILAVQCSVHVI-ANGYKISGKEAPYFYNDYTFPKYSGFTLKSSQIRCKADNTWD 651

Db KLSILAVQCSVHVI-ANGYKISGKEAPYFYNDYTFPKYSGFTLKSSQIRCKADNTWD 2726

QY 2722 SQPEGVVLSGQNP-----GTPSNARVYFSDGLVSSSIVVECRGGYATGLSHHCSVNG 2726

Db SQPEGVVLSGQNP-----GTPSNARVYFSDGLVSSSIVVECRGGYATGLSHHCSVNG 707

QY 652 ELPVCEK-GCOPPPGLHGHGHTGNTVF--FVSGMTVDYTCDDGYLLVNGKSHHCMPSG 707

Db ELPVCEK-GCOPPPGLHGHGHTGNTVF--FVSGMTVDYTCDDGYLLVNGKSHHCMPSG 759

QY 2777 TWTSQDEP---CLVINGCDPGLIPANGIRLGNDRFYKTVTYQCVPGMMESHVSVLSC 2832

Db TWTSQDEP---CLVINGCDPGLIPANGIRLGNDRFYKTVTYQCVPGMMESHVSVLSC 819

QY 708 NMSPSAPRCBETCOHVAGSLQELPA-GSRV-----ELVNTSQDQYQLGHAHYQMGOD 759

Db NMSPSAPRCBETCOHVAGSLQELPA-GSRV-----ELVNTSQDQYQLGHAHYQMGOD 2889

QY 2833 TKDITNNGTGYPCALMCKRPPPLIPNGKVVG--SDFMGSSVTVYACLEGYQLSLPAVFT 2889

Db TKDITNNGTGYPCALMCKRPPPLIPNGKVVG--SDFMGSSVTVYACLEGYQLSLPAVFT 819

QY 760 AENGIWFKKXIPLCVYHCHPPEPVVYNGKHGTGMAENLPGNEVSYECDDQFYLLGEXKQ 819

Db AENGIWFKKXIPLCVYHCHPPEPVVYNGKHGTGMAENLPGNEVSYECDDQFYLLGEXKQ 2939

QY 2890 C-----EGNGSWTGBLPCCF---PVF-CGDPGVPS--RGREDGFGFVRSVSFSCHPPLV 2939

Db C-----EGNGSWTGBLPCCF---PVF-CGDPGVPS--RGREDGFGFVRSVSFSCHPPLV 879

QY 820 CRSDSKHGSGNSGSPQCLNSPPTVTRCPNPEVKGHYKLVKTHSAYSHNDIVVYDCNPGFI 879

Db CRSDSKHGSGNSGSPQCLNSPPTVTRCPNPEVKGHYKLVKTHSAYSHNDIVVYDCNPGFI

[illegible]

```

QY 1443 LLSLSGSHTEBSLPLATSNQVLTKFSAKGLAPARGFFHVQAVPRTSATQCCSSPEPREY 1502
DB 779 RL-----ISSRK- 785
QY 1503 GKRLGSDPSGAIYRFECNSGYALQGSPEIECLPVPALACQNVSAPTCVV-----PCCGN 1558
DB 786 -----FPVQATVQYVDDQGFVLTGSAIILCHDRQAGSPKMSDRAPRCLLEQFRPC--- 835
QY 1559 LTRRGITLSGFPPEPYLINSICWAKIVPEGAGIQIQVVSFTVEQWMSLVEVPDADNT 1618
DB 836 ----- 835
QY 1619 VTMLGFSFGTTVPALNLNSTNQVLHFPYSDISVSAAQGHLEKTYGLSSCCEPAVPSGV 1678
DB 836 -----H-----GLSA-----PENGA 845
QY 1679 KTGERYL--VNDVVSFOCEPGYALQGHANHSICMGPTVRWNVPPPLCTIA 1725
DB 846 RSPERLHPAGATITHFSCAPGYLVKQQAISICVPGHPSHWDPPICPA 894

RESULT 8
149540
procollagen C-endopeptidase (EC 3.4.24.19) precursor, collagen-like splice form - mouse
C.Species: Mus musculus (house mouse)
C.Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C.Accession: I49540
R.Fukagawa, M.; Noboru, S.; Hogan, B.L.M.; Jones, C.M.
Dev. Biol. 163, 175-183, 1994
A.Title: Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1) which is re
A.Reference number: I49540; MUID:94422342; PMID:8174772
A.Accession: I49540
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-991 <RES>
A.Cross-references: UNIPROT:P98063; GB:L24755; NID:g4339606; PIDN:AAA7306.1; PID:g433960
C.Genetics:
A.Gene: Bmp-1
C.Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology;
C.Keywords: hydrolase; metalloproteinase; zinc
F.135-326/Domain: astacin homology <AST>
F.556-592/Domain: EGF homology <EG1>
F.596-705/Domain: C1r/C1s repeat homology <C1R>
F.712-747/Domain: EGF homology <EG2>
F.712,722,728,777/Binding site: zinc (His, His, His, Tyr) #status predicted
F.219/Active site: Glu #status predicted

Query Match 4.4%; Score 755; DB 2; Length 991;
Best Local Similarity 25.2%; Pred. No. 9,66-35;
Matches 235; Conservative 126; Mismatches 338; Indels 232; Gaps 29;

QY 952 ENTSGFELHFSFELIKCEDPGTPKGYKVDGHPAAGSVS-----FSCDPGYSLRGS 1006
DB 244 ENIQPGGEYNFLKMEQVEVESIGET---YDFDSIMHYARNTPFSRGIFLDTIVPYEVANGV 300
QY 1007 EELLCLSGERTMD-----RPLPTCAEBCGTVRGVSGQVLSPCGPAHYENLNINIM 1059
DB 301 KPSI---GQRTLRSLKGDIAQARKLYKCPA-CGETLQ-DSTGNFSSPEYPNGYSAMHRCVW 355
QY 1060 TIEAAGCTTGLHFLVPTBEV---HDVLRIMGPRESSGVTLRELGGPALPKDLHSTFN 1115
DB 356 RISTVPGKILNTSMDLNYSRLCWYDVEVRGPFARKVWVRFGCGKLPKPIVSTDS 415
QY 1116 SVTLQFTSDPFTSKQGFALQPSVSTASTSCNDPGIPONGSRSGDSWEA-GDSTVPQCDBPY 1174
DB 416 RLWVEF-----RSSSNWVGKGFPAVVE----- 437
QY 1175 ALQGSABISCVKIENRFWQSPPTCIAPCGGDLTGSPGVLSINYPPEPYPGKECMKY 1234
DB 438 -----ALCGGVKKDKNGHIGQISFNYPDYRPSKCYKTRI 470
QY 1235 TVSPDYIALVFNINLEP---GYDFLHYIDGRDLSPLIGSEFGSGLPGRISSSNSL 1290

```

```

Db      471 QVSEGFHVGTLFQSFSEIRHDSCAVDYLEVRDGHSESNLIGRCVGENPDIDISTSRLL 530
Qy      1221 FLAFPSDASVSNAGVVIDYTNENPRESCPDPSINKGRVSGDLKLGSSVTTY---CHGGY 1347
Db      531 WLKFSVDSISNAGAVNPFKEVDE-CSR-----NNGGGEORCLNTLGSYKSCDPBGY 583
Qy      1348 EVEGTSTLSCILGPDGKPVMMNPPVCTAPCGGQVYSGDVVLSPNYPONTTSGQICLYF 1407
Db      584 E-----LAPDKR-----CEACGGFLTKLNGSITSPGMPKEPPNNKNTIWO 625
Qy      1408 VTPKDYVVFQGFAPFHTALNDV---VEVHDGHSQHSRLSLSGSHGESLPLATSN 1462
Db      626 LVAPFYRISLQDFEFTEGNDVCKYDFEVRSGILTADSKLHGKFCSEKREVI-TSQYN 684
Qy      1463 QVLKFSKAGLAPARGF--HFVYQAVPRTSATQSSVPEPRYGRKLSGDSVGAIVAFEC 1520
Db      685 NMRVEFSKDNVSKKGFKAHF-----SDKDECS---KONGGQODCVNTPFGS-YECQC 734
Qy      1521 NSGYALOGSPRIECLPVFGALAQWNVSAPTCVPCGGLTERRGTLISPGPEPYLSLN 1580
Db      735 RSGFVLHDKNH-DCKE-----ACGCHKVSTSGTITSPNMPDKYPSKE 777
Qy      1581 CVMKIIVPEAGIQTQVSVFTEQN---WDSLEVPFGADNTVTMLGSEGTVPALLNS 1636
Db      778 CTMAISSTPGHRVXLTLEMNDIESQPCAYDHLFVFGDRAKAPVLCRFGSKKPEVLA 837
Qy      1637 TSNOLYLHFYSDISVSAAGFHLKTYVGLSCEPEAVPSNGVKTGERIYLVNDVVSFCPE 1696
Db      838 TGNMFLRFYSDNSVQKRGQASHST----- 863
Qy      1697 GYALOGHAHISCMFGTVRRMNYPPPLCIAOCGGTVE---EMEGVILSPGF-PGNYPSNMD 1752
Db      864 -----ECGGQVRAADVKTDLXSHAGDNNVPGVD 894
Qy      1753 CSMKIALPVPGGAHIOFLNF---STEPNHDYIEIRNGPYETSRMRMGREGSLPSLS 1808
Db      895 CEMVIYAEQYGVLELVQTREVEBEETDGYLELFGDYSTAPRLGRYCGSGPPEVYS 954
Qy      1809 TSHETTVYFHSQNRPGFKLEYQAYEQE 1839
Db      955 AGDSVLVFKFHSDDTISKGFHLRTYTKPOD 985

RESULT 9
B58788
Procollagen C-endopeptidase (EC 3.4.24.19) precursor, tolloid-like splice form - human
N/Alternate names: bone morphogenic protein 1, tolloid-like splice form
C/Species: Homo sapiens (man)
C/Date: 28-Mar-1998 #sequence_revision 09-Apr-1998 #text_change 09-Jul-2004
C/Accession: AJ37278; B58788
R/Wozney, J.W.; Rosen, V.; Celeste, A.J.; Mltsock, L.M.; Whitters, M.J.; Kriz, R.W.; Hew
Science 242, 1528-1534, 1998
A/Title: Novel regulators of bone formation: molecular clones and activities.
A/Reference number: A37278; PMID:89072730; PMID:3201241
A/Accession: A37278
A/Molecule type: mRNA
A/Residues: 1702; EKRPAIQPRGRPHQKRVQKRNTPQ<WOZ>
A/Cross-references: GB:135279; NID:9619860; PIDN:AA641710.1; PID:9619861
A/Genes: GDB: BMP1; BMP-1
A/Cross-references: GDB:125203; OMIM:112264
A/Map position: 8p21-8p21
C/Function:
A/Description: catalyzes hydrolysis of the carboxyl-terminal propeptide of collagen type

```

```

C/Superfamily: procollagen C-endopeptidase; astracin homology; C1r/C1s repeat homology;
C/Keywords: alternative splicing; beta-hydroxyasparagine; bone; calcium; duplication; g
F/1-22/Domain: signal sequence #status predicted <Sig>
F/23-986/Product: procollagen C-endopeptidase tolloid-like splice form #status predicted
F/110-321/Domain: astracin homology <AST>
F/322-431/Domain: C1r/C1s repeat homology <C1R>
F/435-544/Domain: C1r/C1s repeat homology <C1R2>
F/551-587/Domain: RGF homology <EG1>
F/581-700/Domain: C1r/C1s repeat homology <C1R3>
F/707-742/Domain: RGF homology <EG2>
F/747-856/Domain: C1r/C1s repeat homology <C1R4>
F/860-973/Domain: C1r/C1s repeat homology <C1R5>
F/91-142,332,363,599/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/163-319,185-205,322-348,375-397,435-461,488-510,551-553,559-572,574-587,591-617,644-6
F/213,217,223,272/Binding site: zinc (His, His, His, Tyr) #status predicted
F/214/Active site: Glu #status predicted
F/565,720/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted

Query Match 4.4%; Score 746; DB 1; Length 966;
Best local similarity 25.0%; Pred No. 3,1e-34;
Matches 233; Conservative 125; Mismatches 341; Indels 232; Gaps 29;

Qy      952 EYTSKGFELHFSFELIKCEDPTPKFGYKVDGHPAGSSVS-----FSCDPGYSLRGS 1006
Db      239 ENIQGOEYVFLKMPQEVESLGET---YDFSIHVAHNTSRGIFLDITVPKYEANGV 295
Qy      1007 EBLCLSGRRTW-----RPLPTVAEGCGTVRGEVSGVLSFGYPAPYEHNLCTW 1059
Db      296 KPPT---GQRTRLSKGDIQAARKLYKCPA-CETLIQ-DSTGNFSSPEYNGNSAMHGW 350
Qy      1060 TIEAAGCTIGLHFLVPTVEV---HDVLRIWDPVSSGVLLKELSPALPKDHLSTFN 1115
Db      351 RISVTPGEKILINFSLDLNYSRLCWDYVEVRDPMKAPLRGRFCSKLEPIVSDS 410
Qy      1116 SVTLQFSTDFTSKQGPALIQSVSTATSCNDPGLPONGSRGDSMEA-GDSTVFQCDPGY 1174
Db      411 RLWVEF-----RSSSNVWGKGFAYVE----- 432
Qy      1175 ALQGSABISCVKIENRFMPQSPPTCIAPCGGDLTGPSGVLLSPVPEPYPRGKCDKV 1234
Db      433 -----AICGDVKKDYGHIQSPNYPDYVPSKCIWRI 465
Qy      1235 TVSPDYVALVFNIFNLEP---GYDPLHIYDGRDLSPLIGSFYQSOLPRIESSNSL 1290
Db      466 QVSEGFHVGTLFQSFSEIRHDSCAVDYLEVRDGHSESNLIGRCVGENPDIDISTSRLL 525
Qy      1291 FLAFPSDASVSNAGVVIDYTNENPRESCPDPSINKGRVSGDLKLGSSVTTY---CHGGY 1347
Db      526 WLKFSVDSISNAGAVNPFKEVDE-CSR-----NNGGGEORCLNTLGSYKSCDPBGY 578
Qy      1348 EVEGTSTLSCILGPDGKPVMMNPPVCTAPCGGQVYSGDVVLSPNYPONTTSGQICLYF 1407
Db      579 E-----LAPDKR-----CEACGGFLTKLNGSITSPGMPKEPPNNKNTIWO 620
Qy      1408 VTPKDYVVFQGFAPFHTALNDV---VEVHDGHSQHSRLSLSGSHGESLPLATSN 1462
Db      621 LVAPFYRISLQDFEFTEGNDVCKYDFEVRSGILTADSKLHGKFCSEKREVI-TSQYN 679
Qy      1463 QVLKFSKAGLAPARGF--HFVYQAVPRTSATQSSVPEPRYGRKLSGDSVGAIVAFEC 1520
Db      680 NMRVEFSKDNVSKKGFKAHF-----SDKDECS---KONGGQODCVNTPFGS-YECQC 729
Qy      1521 NSGYALOGSPRIECLPVFGALAQWNVSAPTCVPCGGLTERRGTLISPGPEPYLSLN 1580
Db      730 RSGFVLHDKNH-DCKE-----ACGCHKVSTSGTITSPNMPDKYPSKE 772
Qy      1581 CVMKIIVPEAGIQTQVSVFTEQN---WDSLEVPFGADNTVTMLGSEGTVPALLNS 1636
Db      773 CTMAISSTPGHRVXLTLEMNDIESQPCAYDHLFVFGDRAKAPVLCRFGSKKPEVLA 832
Qy      1637 TSNOLYLHFYSDISVSAAGFHLKTYVGLSCEPEAVPSNGVKTGERIYLVNDVVSFCPE 1696
Db      833 TGNMFLRFYSDNSVQKRGQASHST----- 858

```

1697 GYALQGHAIHSCMGTYRRMYPPPLCAAGCGTYE---EMEGVILSRGF-PGNYPSNMD 1752
859 -----ECGGGQVADVKYKDYLSHAQFDNNYPCGVVD 889
1753 CSWKALPVGCGAHIOFLNF---STEPNDYIEIRNGPYTSHMGFGSSSELPSSLLS 1808
890 CEWIVAESEGYVAVLFTQTFEVEBETDCGYDMLFCQYDSTALSLRGYCSGSPREYVS 949
1809 TSHETTYFHSDSQNRPGFLLEYQAYELQE 1839
950 AGDSVLVWFHSDDTYTKGFLRYTSTFKQD 980
RESULT 10
NBH0H
complement factor H precursor, long splice form [validated] - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text change 09-Jul-2004
C/Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298
R/Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.
Biochem. J. 249, 593-602, 1988
A/Title: The complete amino acid sequence of human complement factor H.
A/Reference number: S00254; MUID:88134059; PMID:2963625
A/Accession: S00254
A/Molecule type: mRNA
A/Residues: 1-1231 <RIP>
A/Cross-references: UNIPROT:P08603; EMBL:X00716; NID:931964; PIDN:CAA68704.1; PID:931965
A/Note: 402-Tyr was also found
A/Note: parts of this sequence, including the amino and carboxyl ends of the mature prot
R/Estaller, C.; Schwaeble, W.; Dietrich, M.; Weiss, E.H.
Eur. J. Immunol. 21, 799-802, 1991
A/Title: Human complement factor H: two factor H proteins are derived from alternatively
A/Reference number: A60238; MUID:91184292; PMID:1826264
A/Accession: A60238
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-56,117-1231 <EST>
A/Note: only portions of this 4.3 kilobase mRNA were sequenced
R/Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.
Biochem. Rep. 7, 201-207, 1987
A/Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human comple
A/Reference number: A54726; MUID:88025472; PMID:2889480
A/Accession: A54726
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-579,1231 <DAY>
A/Cross-references: GB:MI7517; NID:9180497; PIDN:AAA52016.1; PID:9180498
A/Note: parts of this sequence were determined by protein sequencing
R/Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.
Biochem. Rep. 6, 65-72, 1986
A/Title: Partial characterization of human complement factor H by protein and cDNA sequ
A/Reference number: A61565; MUID:86188123; PMID:2938641
A/Accession: A61565
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-1050-1057, 'T', 1059-1102 <R12>
R/Sim, R.B.; Discolo, R.G.
Biochem. J. 205, 285-293, 1982
A/Title: Purification and structural studies on the complement-system control protein be
A/Reference number: A26505; MUID:83048213; PMID:6215918
A/Accession: A26505
A/Molecule type: protein
A/Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>
R/Bartlow, P.N.; Norman, D.G.; Steinkaseker, A.; Horne, T.J.; Pearce, J.; Driscoll, P.C.
Biochemistry 31, 3626-3634, 1992
A/Title: Solution structure of the fifth repeat of factor H: A second example of the com
A/Reference number: A44551; MUID:92232849; PMID:1533152
A/Contents: annotation; NMR structure determination; residues 264-292
R/Norman, D.G.; Bartlow, P.N.; Barton, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.
J. Mol. Biol. 219, 717-725, 1991
A/Title: Three-dimensional structure of a complement control protein module in solution.
A/Reference number: A49224; MUID:91278097; PMID:1829116

A/Contents: annotation; NMR structure determination; residues 927-985
R/Estaller, C.; Koerfgen, V.; Schwaeble, W.; Dietrich, M.P.; Weiss, E.H.
J. Immunol. 146, 3190-3196, 1991
A/Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a nove
A/Reference number: I56100; MUID:91201892; PMID:1826708
A/Accession: I72654
A/Status: translated from GB/EMBL/DDB
A/Molecule type: mRNA
A/Residues: 1047-1231 <RES>
A/Cross-references: GB:M65294; NID:9183766; PIDN:AAA35948.1; PID:9183767
R/Carton, J.A.; Bates, R.C.; Smith, A.I.; Tetz, T.; Arellano, A.; Gordon, D.L.; Burns,
Biochim. Biophys. Acta 1289, 305-311, 1996
A/Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.
A/Reference number: S66298; MUID:96205365; PMID:8620012
A/Accession: S66298
A/Status: preliminary
A/Molecule type: protein
A/Residues: 411-419,574-578,580-582 <CAR>
C/Comment: Factor H has also been found bound to cell membranes in an unknown manner. H
C/Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver
C/Genetics: <HF1>
A/Gene: GDB:HF1; HF
A/Cross-references: GDB:120041; OMIM:134370
A/Map position: 1q32-1q32
C/Genetics: <HF2>
A/Gene: GDB:HF2; HF
A/Cross-references: GDB:129095
A/Map position: 1q32-1q32
A/Note: the correspondence between the two loci and the sequences indicated is unclear;
C/Function:
A/Description: a cofactor in the inactivation of C3b by serine proteinase I; also incre
he alternative complement pathway
A/Pathway: complement alternate pathway
C/Superfamily: complement factor H; complement factor H repeat homology
C/Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1229/Product: complement factor H #status experimental <MP>
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>
F:85-141/Domain: complement factor H repeat homology <FH01>
F:146-205/Domain: complement factor H repeat homology <FH02>
F:210-262/Domain: complement factor H repeat homology <FH03>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:566-623/Domain: complement factor H repeat homology <FH10>
F:630-684/Domain: complement factor H repeat homology <FH11>
F:691-744/Domain: complement factor H repeat homology <FH12>
F:753-803/Domain: complement factor H repeat homology <FH13>
F:811-864/Domain: complement factor H repeat homology <FH14>
F:870-926/Domain: complement factor H repeat homology <FH15>
F:931-984/Domain: complement factor H repeat homology <FH16>
F:989-1043/Domain: complement factor H repeat homology <FH17>
F:1048-1102/Domain: complement factor H repeat homology <FH18>
F:1109-1163/Domain: complement factor H repeat homology <FH19>
F:1167-1228/Domain: complement factor H repeat homology <FH20>
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,35
1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent
F:529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 4.2%; Score 721; DB 1; Length 1231;
Best Local Similarity 23.1%; Pred. No. 1,1e-31;
Matches 314; Conservative 192; Mismatches 516; Indels 340; Gaps 81;
1838 QECDEPFPAN-GIVRA---GYNQGSYTFECLPGYOLGHPVLTCQGTNRNMD--H 1890
19 EDCELPFRNTTELTSWSDQYPECTQAIYKRPYRSGNVIWCRG---EWALN 75


```

Db 2036 ----IRVYGLKLEK-----DKFSFTWDPV--PLNYRNLTLAKD-----KXAKYDCC--- 2075
Qy 341 FTSSGCVLSPNYED-----YGNHLHCWLLIAREPSRIHLAFNDIDYEP- 386
Db 2076 ----GGDLFTPEYEMDYFTNPLPAGOSFGYEENLHCWIIIRPMFTGIELFEDYLEBDV 2130
Qy 387 ----QFDFLVIR-----DGAFAEAPVLGTSGNQLPSISSSGVARLEQTD 430
Db 2131 QNCAPDFVSRLQFDDOPEENDVDVLSVSKCALVSNNTFNFSVNRALHI--HFVTD 2187
Qy 431 HSTCKRGF-----NITFTFRH----- 447
Db 2188 RSRHIGFKLKYRLTCLNSFEHIRGIFREHTLTSPNYDGNHTPSWKCOYSLIIBNRK 2247
Qy 448 -----NECPDPGVVNGKRPD-----SLQSSSISPLCDE 478
Db 2248 VSVEIUDLDIOENSPCSYNNALLIGNRPSBELVETHSOKSKYCKLEQGERLNFTSAR 2307
Qy 479 GFL-----GTQSEETITCVLEKGSVYVNSAVLRCEAPCGGHLTSPSGTILSPGMPRPYK 532
Db 2308 GRLLTKYNSGPNKSGKGFMMITKEQUTECPSGVLHVDD-----NSPSKVLSPFPFORIP 2361
Qy 533 DALSCAWIEAOPGYPIKITPDRFKTEVN-----YDLEVRDGRGTYSAPLIGVYHG 584
Db 2362 NSVCEVYMAAPNGHRLMLTFDSENFDIDGNQKODLDYIEIRDGPVHSEVIGYICGN 2421
Qy 585 QVPOFLISTSNYLYLLFSTDSKSHDIFOLRYETITLQSDHCLDPGIPVNGQRHNDPV 644
Db 2422 QAPSTIFSTNYFLYMKLHT-----SEYKSRRF----- 2449
Qy 645 GALVTFSCDSGYTISDGEPLCEBNFOMKRLBPCEALCGFI---QSSGTILSPGRPD 701
Db 2450 --VATYGI-----ATCGGTVMREVVTHITSPSPY 2479
Qy 702 FYPNNLNCTWIIETSHGKGVFTFTHLESCHDYLITENGSTFQPLQOLGSRLPAP 761
Db 2480 PFAPVQCGQWVKNP-----THMIKAKYDHWLFPYNPCTMEQLMIR----- 2522
Qy 762 SAGLYNFTAOVRFISDFSMSYEGFNITFSEYDLEPCEPE--VPASIRKGLQGVGDTL 820
Db 2523 ----DGNSTA-----NPLIGPVCVPRHA-----PDVF 2545
Qy 821 TFGCFPGVRLGEGTARITCLGRRRLMSPLRC---VA-----ECGNSVTGOTGLSP 871
Db 2546 TRSASNOITVQFTNSITTRGRCQYCSN--KKGCFDAVATNSGKCGGRITNLGLTLP 2603
Qy 872 NFPVNNNNHBCIYSIQTPGKGIOLK---ARAFELSEGDVLKV----- 912
Db 2604 GYPRKLLSHVYKCIDFKMKRPFITRFKITTPGANAAYHKFDNQVRIAKQHFPPVAVYEG 2663
Qy 913 ---YDGNNSA-----RLIGVFSHS--EMMGVTLNS--TSSSLWLDPITD 950
Db 2664 FPPYDGAQVSTYFCKNETQITSITDTIRIYVDSHTEMISAINDDQOLNSTFAPFTID 2723
Qy 951 ----AETSNGFELHNS-----SFLKCEDPGRPK-----F 978
Db 2724 YMAIPANSENGCTLIQINKLITTEPHLANKVNGKGVGLDSFCHVSIKPPANFESVHLEIS 2783
Qy 979 GYKUHDSGHFAGSVSVSFSCDPSYSLRGSEELLCLSGERRTWRDPLPTCVAB----- 1029
Db 2784 DYKADSELOPAGAK--CFBGMGHLKIKSDPY-----PIETTVCDGKMKRQOS 2830
Qy 1030 -----CGGTVRGVSQVLSPPGYA--PYEHN 1054
Db 2831 EMIVNPQIDLFISQLFRDQEQFNLITIFQCKGCVISSPNTDITSPNGSGIKLPLG 2890
Qy 1055 LNCIWTIEAEGCTIGLHF---LVPTREYHVDLRIWDPVESGVLLKELSGALPKDL 1110
Db 2891 SKKCMVLEAPBGQIVKXIVEMQIAYHDECENHLLIGEG-----ROADVNTIHKYC 2942
Qy 1111 HSTNSVVLQFSTQFTSCKQFALQFSVSTATSCNDPQIPONGSRGDSW-----EAG 1163

```

```

Db 2943 HKMDGE-----OEOKLEDRFKI-----IKSHGRYLTLIWTNMKYEABG 2981
Qy 1164 DSTVFOCDPGVALOGSAHEISCVKIENTFPMQSPPTCIACGGDLTGPSGVILSPN-PE 1222
Db 2982 -----WKIQVEFLNENE-----BCGHTSGMSGTITYPMGDK 3014
Qy 1223 PYPGKECDMKWTVSPDYIALVFNILFNL-----PGYDFLHI 1260
Db 3015 DYENDELECWMDIQVPLGYHNLKFKDPDVTAEKACADQOLLIOESHSTRANSRPGDYFL 3074
Qy 1261 YDGRDSLPLIGSFYGSQLPGRLESSNSLFLAFRSDASVSNAGFVIDYENPRESCFDP 1320
Db 3075 FODEKETPLC---GIEHPKDFESBSNRVRLNFTTDSQTAGFRVNW----- 3119
Qy 1321 GSIRKGRVSDLKLGSSVTVYCHGYEVEGTSTLSCLLPDQKPVNNRPNVCTARCG 1380
Db 3120 -----EABCGA 3125
Qy 1381 QYVSGDGVLSPNYPONYTSGQICLYFVTVP--KDYVFGQAFAPHTALNDVVEVHDSHQ 1439
Db 3126 IYRLNHGVTISPYPNQYPRNDITCTYLIAPPDQNSVLAIRFADP----- 3169
Qy 1440 HSRLLSLSGSHGSGESLPLATSNQVLIKFSAKGLAPARG--FHFVQAVERTSAQCSSV 1497
Db 3170 -----DLATMR-----TSFGAPCEDDYLIQIDGTDRVLHTPCASE 3206
Qy 1498 PEPR-----YKRLGSPSYGALVRFECNSGYALQSPBELCLFVPGALQMNVSAP 1549
Db 3207 PMPKDALVFKAIGLKFKVSDKS--YYMEEDNKGKGS-----KCPOLAYSIN-- 3253
Qy 1550 TCVPVPCGNLTERG-----TILSPGPEEYLNLSNCVMKIYVPEGAGIOIUVVSFVTEQ 1604
Db 3254 ---KCGDNLEBEGTFITTVTSPAPFLYTKDLCVMMITTDPNQRLNIREMMLA 3309
Qy 1605 ---NMDSLEVPDQADTVT--MLGSFSGT--TVPALINSTN--QLYLHFYSDISVSAAFP 1656
Db 3310 FHDCSADYIEFPDSSDIMAKTLGKFCGTMDKIPQYRIVTSGPVLILHMRDFFVNGGF 3369
Qy 1657 HLEKTYGLSSCPEPAVPSNGVKTGERYLVNDVVSFQCEFGYALQGHANISCMPTVRW 1716
Db 3370 KL-----AVISLQKQEG----- 3382
Qy 1717 NYPRPLCIAQCGG--TVBMEGVILSP--GPPGNVPSNMDCSWKIALPVGGAHIOFLNFS 1773
Db 3383 -----CGGLTITTSWQTLTSPKXEDDGNIPRALLCGWITSGVVDQOLIRIDGV 3442
Qy 1774 TE-----PNHDYLE--IRNGPYETSRMMGR--FSGSELPSLSLSTSHETTVYFH 1818
Db 3433 TEOLEYPRGDRPSPECIDALAIYDQGFEPPLAAGDICTGTRPKLTYTSHRAFTFE 3492
Qy 1819 SDHSQNRPGFLEYQAYEL-----OEC----- 1840
Db 3493 TDRDGTGRGFNISISYIDSDCGGLKATNEIKTLVYKIGITSDDNKEMKERSHQRCRFMI 3552
Qy 1841 --PDPEPANGIYRGAGVNYGQSVTFECLGQULGHPVLVLCQHTNRMMDHLPKREV- 1897
Db 3553 QGPKTEPV---IYVFOQFNI--PSAGDCSPSFV--EIIYSADIMPKNSRSLKRYN 3605
Qy 1898 ---PCGGIN--TSSNGTVVSPGFPSPYSSQDCWMLITVPIGHVNLNLSLQTEPSGD 1951
Db 3606 ILDSNRTIDNTIKSRLTSPNVPQVYSNSTCV----- 3640
Qy 1952 FITIWDGPQOTAPRLGVFTRSMAKTYQSSNOVLLKFRHDAATGIFAIFSAVPLTKC 2011
Db 3641 -----TWLQSSNOKML-----LVFNDFTLBE- 3661
Qy 2012 PPPTILPNAEVTEBERFNI-----GDI--VRYPC---LPQFTLVGNELLCCKGTVYQFE 2062
Db 3662 -----PNN--VNNKCDYDYLMLKEGDSNGTYCGSLTLPALMTSGKOLFTALLKSHSLN 3713
Qy 2063 GPPPICVHCPTVB-----LLTDS--TGVLISQSYGSYPOFOTCSMLVR----- 2105
Db 3714 HAGYDASVYTVVSRDQIQPADSYELGYISSIGYNGVINKSVTQFTLIRPNSHDCSI 3773

```


QY 2106 VEPDNYSLT---VEYFLSEKQYDEFEIFDGPSSGSLTKA-----LSGNYSAPLVT 2156
 Db 3774 IFTDVISMTQLPBECFCFPTSEYIEYV---KRTQVKARIRSCFPAKKAELILEA 3829
 QY 2157 SSNSVYLKMS---SDHAYNRKGFIRY 2180
 Db 3830 DNNDRYIKFAFKSAKSENDGRGFIRY 3857

RESULT 12

A45900
 Complement C3b receptor type 2 long form precursor - mouse (fragment)
 C/Species: Mus musculus (house mouse)
 C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
 C/Accession: A45900; 148306
 R/Kurtz, C.B.; O'Toole, B.; Christensen, S.M.; Weis, J.H.
 J. Immunol. 144, 3581-3591, 1990
 A/Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene
 A/Reference number: A45900; MUID:90229754; PMID:2139460
 A/Accession: A45900
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-676 <KUR>
 A/Cross-references: UNIPROT:Q99254; GB:M36470
 A/Experimental source: Clone 31-1
 R/Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.
 J. Exp. Med. 181, 151-159, 1995
 A/Title: Mouse complement regulatory protein Crty/p65 uses the specific mechanisms of bc
 A/Reference number: 148306; MUID:9510691; PMID:7528766
 A/Accession: 148306
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 21-367 <RES>
 A/Cross-references: EMBL:U17128; NID:g595980; PIDN:AA178271.1; PID:g595982
 C/Genetics:
 A/Gene: Cr2
 A/Introns: 80/1, 113/2, 142/1, 275/1, 333/1
 F/22-78/Domain: complement factor H repeat homology <FH01>
 F/83-140/Domain: complement factor H repeat homology <FH02>
 F/145-211/Domain: complement factor H repeat homology <FH03>
 F/217-272/Domain: complement factor H repeat homology <FH04>
 F/276-331/Domain: complement factor H repeat homology <FH05>
 F/336-394/Domain: complement factor H repeat homology <FH06>
 F/399-458/Domain: complement factor H repeat homology <FH07>
 F/467-523/Domain: complement factor H repeat homology <FH08>
 F/531-587/Domain: complement factor H repeat homology <FH09>
 F/592-648/Domain: complement factor H repeat homology <FH10>

Query Match 4.0%; Score 677; DB 2; Length 676;
 Best Local Similarity 26.1%; Pred. No. 1.5e-30;
 Matches 209; Conservative 95; Mismatches 300; Indels 198; Gaps 33;
 QY 2130 IFDGPSSGSLTKA LSGNYSAPLVTSSN-SVYLKMS---SDHAYNRKGFIRY 2180
 Db 14 IFTGVIGCKLTPKYS--FAKESIVDKSEFAIGTWYKCRPGFRKSFITTCETSKW 71
 QY 2181 SADV-----CSLPAPLHGFLLQGTSTQPGSIFHFGNAGYRLVGHSMALCTRHPOY 2233
 Db 72 SDAQCFKRRKPCMNPPQELHGSVHINTGIEFGSTIYSCNQGRIIDGSSACIVSDMTV 131
 QY 2234 HLMSEAPICQALSCGLPEAPKNGWFGK---EYVGTAVYSCSEGHILQ-----AGAE 2285
 Db 132 -MNDNDMPCESTPCSSPAISNGDFYSSRSFEGVAVTYVYCHTKKREKLPDLVGEK 190
 QY 2286 A---TAECLDTGLMNRNVPQCPV-VTCPDVSSISVEHGRWRLIFETGYQQAQMLMIC 2341
 Db 191 SYCTSKDNQGVIMN--SPPCQIRVYKCP---MEIENG-----LV 227
 QY 2342 DGGYVYTGGRVRCQANGKWSJGDSPTTCRIISCEGLPIPNGHRIGTLVYGATAIRFSC 2401
 Db 228 ESGFKHS-----FPLND-----TVIFKC 245

QY 2402 NSGYTVGSRYRECMANGLMSSGEVRCCLAGHCTPEPIYNGHIN--GENYSYRGSVYQC 2459
 Db 246 KSGFTWKGSRIRAWCQPSKMSSEPLPTCFMG--CLPQNIHGHYKNDFFSVQGVKSYTC 304
 QY 2460 NAGFRLIGSVRLICQDDHHMSGKTPCVPIITGCHPQNPVNGLTQGNQF-----NLNDYK 2514
 Db 305 NRGYTLIGNLVECTSLGTSWNTVPCEVYKSC--DAIP-NHLHGRVFLPMLQGAEVS 361
 QY 2515 FVCNPGYMAEGARSGCLASGO---WSDMLPTCRILINCTPGHNSVYQVAGSPHRS 2571
 Db 362 FVCDLGFOLKGRSSQCIPEGETVIMNNKFPVCEQISC--DPPRYVNAKRPYSLP--IV 418
 QY 2572 FGTYSYRNHGFYLLGTPVLSG---QGDGWDPRPQC---LLVSGHFGSPHSGM 2623
 Db 419 FGTVALRYTSPSYRLIGETALFICISENVHATWDAAPRCESVNTIISCSDEIVPGGFNM 478
 QY 2624 SGDS--YTVGAVAVRYSCIGKRTLVNSTRMCLDGHM--TGSIPHCSGTSVGVCGPGRPA 2680
 Db 479 KSKXAPFRIGDGVITTCANFTMKSKTYWCQANEMGPTALPVESDPFLFC--PSLPT 536
 QY 2681 --HGIRLG---DSPDGTVMRPSCEAGHVLGRSSERTCQANGSWGSGQPECVISCNG 2735
 Db 537 INHGHHTGQHVDFVAGLSVTYSCEPGYLLTGKTKIKCLSGDMGVIPTCKEACGHRG 596
 QY 2736 TFSNARVYFSDGLVYSSSIYVECRPGYATGLSHNCV---NGTWGSDPECLVINGD 2792
 Db 597 KRPNQVKEPISLQVGTIVYSCNBYOLOQOPSSQCVIVBOKAIVT----- 643
 QY 2793 PGIPANGRLGNDPFRYNTVYQCVPGYMESHVSVLCTKDRWTNGTKPVCKALMCKP 2852
 Db 644 -----KKVVCNEILCP 655
 QY 2853 PPLINGKTVGSDPFMGSSVTVY 2874
 Db 656 PPVVRNGSHTS---FSENVPEY 674

RESULT 13

NBM5H
 Complement factor H precursor - mouse
 N/Alternate names: protein beta-1-H
 C/Species: Mus musculus (house mouse)
 C/Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jul-2004
 C/Accession: A26154; 149711; 149728
 R/Kristensen, T.; Tack, B.F.
 Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
 A/Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length.
 A/Reference number: A26154; MUID:86233353; PMID:2940596
 A/Accession: A26154
 A/Molecule type: mRNA
 A/Residues: 1-1234 <KRI>
 A/Cross-references: UNIPROT:P06909; GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g387181
 R/Natsunome-sakai, S.; Nataka, M.; Nataka, M.; Harada, Y.
 J. Immunol. 144, 358-362, 1990
 A/Title: Demonstration of an unusual allelic variation of mouse factor H by the complete
 A/Reference number: 149711; MUID:90111033; PMID:2136885
 A/Accession: 149711
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-18 <RES>
 A/Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729
 R/Munoz-Canoves, P.; Tack, B.F.; Vlk, D.P.
 Biochemistry 28, 9891-9897, 1989
 A/Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma in
 A/Reference number: 149728; MUID:90148935; PMID:2533512
 A/Accession: 149728
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-19 <RES>
 A/Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926
 C/Comment: Two codominant alleles of factor H are present in mice.
 C/Comment: Factor H functions as a cofactor in the inactivation of C3b by serine protease
 (C3 convertase) in the alternative complement pathway.

C/Genetics: 1
 A/Map position: 1
 C/Superfamily: complement factor H; complement factor H repeat homology
 C/Keywords: complement alternate pathway; duplication; glycoprotein; plasma
 F/1-18/Domin: signal sequence #status predicted <SIG>
 F/19-124/Product: complement factor H #status predicted <MPT>
 F/21-80/Domin: complement factor H repeat homology <FH01>
 F/85-141/Domin: complement factor H repeat homology <FH02>
 F/146-205/Domin: complement factor H repeat homology <FH03>
 F/210-262/Domin: complement factor H repeat homology <FH04>
 F/246-248/Region: cell attachment (R-G-D) motif
 F/325-385/Domin: complement factor H repeat homology <FH05>
 F/325-385/Domin: complement factor H repeat homology <FH06>
 F/389-442/Domin: complement factor H repeat homology <FH07>
 F/448-505/Domin: complement factor H repeat homology <FH08>
 F/509-564/Domin: complement factor H repeat homology <FH09>
 F/559-622/Domin: complement factor H repeat homology <FH10>
 F/629-633/Domin: complement factor H repeat homology <FH11>
 F/630-743/Domin: complement factor H repeat homology <FH12>
 F/752-802/Domin: complement factor H repeat homology <FH13>
 F/808-861/Domin: complement factor H repeat homology <FH14>
 F/867-931/Domin: complement factor H repeat homology <FH15>
 F/936-989/Domin: complement factor H repeat homology <FH16>
 F/936-989/Domin: complement factor H repeat homology <FH17>
 F/1053-1107/Domin: complement factor H repeat homology <FH18>
 F/1114-1168/Domin: complement factor H repeat homology <FH19>
 F/1172-1233/Domin: complement factor H repeat homology <FH20>
 F/21-66,52-80,85-129,114-141,146-152,178-205,210-251,257-262,267-309,294-320,325-374,357-
 F/802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-
 F/676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) #status pre

Query March 4.0% Score 676.5; DB 1; Length 1244;
 Best Local Similarity 23.3%; Pred. No. 3.7e-30;
 Matches 325; Conservative 155; Mismatches 478; Indels 437; Gaps 81;

1838 QECDEPPEFANG-IVRGAG---YVNGSVTFECLPGYOLGHHVLTCTOHGTNNM--DH 1890
 19 EDCGPPRESEISELSSWSMQLPBEQTQATYKCRPGRTIGTIVKCKNG--KVAASN 75
 1891 PLPLPC-EVPCCGNITSSNGITVSPGSPSSSODCWLITVPIGHVRLNLSLQTEPS 1949
 76 PSRIKCRKPCG-----HPG-DTPFGS-----FRLAVGSPFEFGAKVYVYCD 116
 1950 GDFITINDPQQTARLGLVFRSMAKKTVOSSNOVL--LKHNDATGIFATAPASAP 2007
 117 G-----YOLGEIDYRCGADGWINDIPLC--E 142
 2008 LTKCPPTILPNAEVLT---ENEENIGDIVRYRCLPGFTLVNENILTCKLGTYLQPEG 2063
 143 VVKCLPTELENGRIVGAAMTDOEYFGQVVRPECSGFKIBHK----- 188
 2064 PPRICEVHCPTELLTSTGVILSOSYPSGYPOFCOSWLV---RVEPDYNISLTVEYF 2119
 189 -----EIHSENGELMSN-----EKPRCEVILCTPPRVENGGINVKKVYK 228
 2120 LSEKQVBEFEPDGPSSQSLTKALSGVNSAPLVTSSNSGVYR---WSDAHVYNNKGF 2176
 229 ENERYHYKCKRGVYKERGADVCTGSSSQPFCEBKRCSPYLLNGIYPRHRIHSDD 288
 2177 KIRYSAPV---CS-----LPPAPL-----HGFIIGOTSTOP----- 2204
 289 EIREECVYGFYRVGTSTVSKCTPRGWIAPVRCTLKPCFEPQPKGRLYEBSLRPNFVS 348
 2205 -GGSIHGCGNAGYLVGHSMALCTRHPOGTHLSEBAPLCOALSCGL---PEAR----- 2254
 349 IGNKYSYKCDNGF-----SPSGYS-W-----DYLRCTAQGWEPFVPCVRKC 389
 2255 -----KNG--WVEKEVTVGTKAVYSCSEGYHLQAGABATACSDTGMSRNVPRQCPV 2307
 390 VFHYVENGDSAYMKVYVQGSGLVYQCSLNGSDG--TWTCTENG-WS---PPPCIR 444
 2308 VTCPDVSSISVENG-----RMRLIETQYQFOALMLICDPGYV--YTQG--RVIRCOA 2357

445 IKTCASADHIDNGFLSESSSIYALNRETSYR-----CKQGVYNTGSISSITCLQ 496
 2358 NGKMS-----ELPPP--NGHRIGTSLVYATAPFSCNSGYTLVGRVR 2413
 497 NG-WSPQPSCKKCDMPFENSITKTRTWPKLNDKLDYBCLVGFENEYKHTGSICTY 555
 2366 -----STPTCRITISCG-----ELPPP--NGHRIGTSLVYATAPFSCNSGYTLVGRVR 2413
 556 YGMSDPTSCYERECVSUTLRLKVLVSPRKKRYVGLD-----LEFCHSGHR-VGDSV 608
 2414 ECNANGIWSGSEVRC--LACHCGTPEIYVNGHNGE---NYSYRGSVYVQCMAGFRLLIG 2468
 609 QCHYFG-WSPQFPCKQOVASCAPLEIILNGEINGAKVYSHGEVVKYCKCPKRFLLKGP 667
 2469 SVRIQCOODHMSGKTPECV--PITCGHPGPNVAGLTOGN-----QNLNDVYKFCVNP 2550
 668 NKIQ-CVDGWT-TLPPICEEERTCG---DIELEHSAKCSVPYHNDVSVEFICEEN 721
 2521 YMAEGARSGCLASGQW-----SDMLPTCRINCTDGHQENSRYOVHASGPH--RF 2570
 722 PTMIGHSSVSCI-SGKWTQLPKCVATDQLEKCYLAKSTG-----IEAIKPKLTER 770
 2571 SFGTWSYRCNHGFTYLLGTPVLSCQGDGTWRDRPOCL-LVSCGHPGSPPHSQ----- 2622
 771 THNSTMDYKCRDKQEVERSICI---NGKWD-DEPNCTSKTSCPPPOIENQVIEETVK 825
 2623 -MSGDSITVAVYRISCTIGKRTLVGNSTRCGLDGHWTGSLPHRCSGTSVGVCGDPGIPAH 2661
 826 YLQBEKSLVLCODNY-----LTQSEENVCKDGRWQ-SLPRC--TEKIPCSQPTIEH 875
 2682 G-IRL-----GDSFDPGTVMRFCEAGHYLVGSSERTQANSGMSGQPECGV 2728
 876 GSINLPSSSERKDSISSHENGTTSYCDGDFRIPERNRITCYA-GKMS-TPPRCVG 933
 2729 ISCGNP-----GTPSNARVVPDGLVFSSSIYVECREGYATGLSRHCSVNGTWTGSDP 2783
 934 LPCGPPSIPILGTVSLSELSYQHG---EBVTYHCSGFGIDGPAFICE-GGKMS-DPP 987
 2784 ECLVINGDGDITANGRLG---NDPRYNTTVYQCYPGTGMESHRSVLSCTKDRTWNG 2840
 988 KCIKTDDVLPVYKNAIIRKSKSYRTGQVFRCSPPQNMGS--DVTQVNSR-WIG 1044
 2841 TKPVCKALMCKPPPLIPNGKV---GSDFMGSSVYACLEGYQLSLPAVTEGNSGWT 2897
 1045 -QPVCKDNCVDPPHPNATIVTRTKKXYLHGRVRYECKKPLBFGQVEVME-NGIWT 1102
 2898 GELPOCFPVECPDGPVPSRGRREDRGRFSYRSVSFSCHPPLVVGSPRRFCQSDGTWGT 2957
 1103 -EKPKC-----RDSTG-----KCGP----- 1117
 2958 QPSCIDPTLTTCADPGVPOFGIONNSQGVGSTVLFRCKGTYLQSTTRTCLPILTWIS 3017
 1118 -PPIDNDITSLSLP-----VYBPLSSVEYQCKYVLLKKGKTIITC-TNGKMS 1163
 3018 GTPPDCCVHHCRQPE 3032
 1164 -BPPTCL-HACVAP 1176

RESULT 14
 A30359
 P-selectin precursor - human
 N/Alternate names: CD62 antigen; granule membrane protein 140
 C/Species: Homo sapiens (man)
 C/Date: 18-Oct-1989 #sequence_revision 30-Sep-1991 #text_change 16-Aug-2002
 C/Accession: A30359
 R/Johnston, G.I.; Cook, R.G.; McDevier, R.P.
 Cell 56, 1033-1044, 1989
 A/Title: Cloning of GMP-140, a granule membrane protein of platelets and endothelium: s
 A/Reference number: A30359; PMID:89168432; PMID:2466574
 A/Molecule type: mRNA
 A/Residues: 1-830 <JOH>

A:Cross-references: GB:M25322
 A:Note: parts of this sequence, including the amino end of the mature protein, were conf
 C:Genetics:
 A:Gene: GDB:SELP; GRMP
 A:Cross-references: GDB:120018; OMIM:173610
 A:Map position: 1q22-1q25
 C:Superfamily: P-selectin; C-type lectin homology; complement factor H repeat homology;
 C:Keywords: cell adhesion; glycoprotein; phosphoribidene; phosphoprotein; surface anti
 F:1-41/Domain: signal sequence #status predicted <SIG>
 F:142-830/Product: P-selectin #status experimental <MAT>
 F:163-194/Domain: EGF homology <EGF>
 F:200-257/Domain: complement factor H repeat homology <FH01>
 F:262-319/Domain: complement factor H repeat homology <FH02>
 F:324-381/Domain: complement factor H repeat homology <FH03>
 F:386-443/Domain: complement factor H repeat homology <FH04>
 F:448-505/Domain: complement factor H repeat homology <FH05>
 F:510-567/Domain: complement factor H repeat homology <FH06>
 F:572-629/Domain: complement factor H repeat homology <FH07>
 F:642-699/Domain: complement factor H repeat homology <FH08>
 F:704-761/Domain: complement factor H repeat homology <FH09>
 F:772-795/Domain: transmembrane #status predicted <TM>
 F:796-830/Domain: intracellular #status predicted <CYT>
 F:54,98,180,212,219,411,460,518,665,716,723,741/Binding site: carbohydrate (asn) (coval

Query Match 3.9%; Score 657.5; DB 2; Length 830;
 Best Local Similarity 24.3%; Pred. No. 2.5e-29;
 Matches 196; Conservative 92; Mismatches 325; Indels 193; Gaps 27;

2294 GLMNRNVPPQ---CYPVCPDVSSIVHGRRLLFFQYQFOAOLMIDCPGYTTG 2349
 143 GKMDHEICLKKKHALCTASQODMS-----SKQGECLFTIGNYCS---CYPGFY--- 190
 2350 QVRIRCOANGKMSLGDSTPCRLI-SCGELPIPN-----GHRIGLSVYGATAIRSCNS 2403
 191 -----GPEVYRECELELEPHVLMNSHPLMNS-FNSQSHCTD 232
 2404 GYTLVSRVRECMANGMSGEVRCLAGHCTPEPIVNGH-----NGENYSYKGSVYVQC 2459
 233 GYQVNGPSKLECLASITWTKPPOCLAACOPKIPERGNMILHSAKAFQSSCSFSC 232
 2460 NAGFRLIGMSVRIQOQHMSGKTPPCVPIITCGHPGPNV-----LTQSGFNLNDV 2512
 293 EBFALVGPVVOCTASGWTAPAPVCKAVOCQHLERSEGTMDCVHPLT--AFAYGSS 349
 2513 VKFVNCNGYMAEGARSQCLASGQWSDMLPTCRIINCTDPGHOENSVRQVHAS-----GP 2567
 350 CKFECQGVYVRGLDMRCIDSGHWSAPLPTCAISC-----EPLESPVHGSMDCSPSL 403
 2568 HRFSGFTVSYRCHNGYLLGTPLYSCQSDGTWDRPPQCLLVSCGHPGSPHSGQSGDS 2627
 404 RAFQYDNTCSFRCAEGFMRLGADIVRCDNIGQWTAPVQALQCODLVPVNEARVN-CS 462
 2628 YTVGAVVRYSCIKRTLVGNSTRMCGLDGHWGSLPHSCGSTSVGVGGDGIPIAHGRLD 2687
 463 HPEGA-FRYQSV-----CS----- 475
 2688 SEPFGVMAFSCAGHVLKSSSERTCOANGSWSGQPECGVISCNPGTPSNARVVFSDG 2747
 476 -----FTCNELLVGASVLOCLATGNMNSVPECOAIPCTPLLPQNGTMTCVPR 526
 2748 L---VFSSSIYVCREGYYATLGLSRHCSVNGTWTSSDEPCLVINGDGIPIANGRLCN 2804
 527 LGSSTSKSTQCFICDGSYLSGPERLDCTRSGRWTDSPMCEAIKPELFAPEQSLDSC 586
 2805 D---FRYNTVYQCVPVGMESHVSVLSTKDRITNGTKPVCKALMCKXPPLIPNGX 2860
 587 DTRGEFNVGSTCHFCNNGFKLEG--PNVVECTTSRMSATPTCTCKGISLFTP----- 638
 2861 VVGSDEMSSVTVYACLEGYQLSLPAVFTCEGNSWTGELPOCFPVFCGDPGVPGRGR 2920
 639 -----GLQCPAL--TTPGGGT-----MYC-----RHH 658
 2921 DRGFSYRSSVSFCHPPLVLVGSPPRFQCSQSDGTWSTQPSCIDPILTTTCADGCVQFGIQ 2980

Db 659 PGTFGNTTCYFCNAGFTLLIGDSTLSCRSPSGMTVTPAC--RAVKSEILVNRPIAM 715
 QY 2981 NNSQ---GYOVGSTVLFRCKGYLLOGSTTRTCLPVLTWSGTRPPDVPVHNCQPERPTA- 3036
 Db 716 NCSNLMGNYSYSGICGHLCEGLANGSQTCAQENSHMSTVPTCOAGRLPTIOEALTYF 775
 QY 3037 ----ANVGLDIPSMGYTLITPARR 3057
 Db 776 GGAVASTIGL----MGTLIALTRK 797

RESULT 15
 T31069
 tolloid-BMP-1 like protein 1 - California sea hare
 N:Alternate names: probable metalloprotease TBL-1
 C:Species: Aplysia californica (California sea hare)
 C:date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #ext_change 09-Jul-2004
 C/accession: T31069
 R:litu, Q.R.; Hattar, S.; Endo, S.; MacPhee, K.; Zhang, H.; Cleary, L.J.; Byrne, J.H.; E
 U. Neurosci. 17, 765-764, 1997
 A:title: A developmental gene (tolloid/BMP-1) is regulated in Aplysia Neurons by treatm
 A/reference number: Z20965; MUID:98007484; PMID:8987797
 A/accession: T31069
 A/status: preliminary; translated from GB/EMBL/DBJ
 A:molecule type: mRNA
 A/residues: 1-1070 <lit>
 A:Cross-references: UNIPROT:P91972; EMBL:U57369; NID:g1899041; PID:g1899042; PIDN:AA474
 C/superfamily: dorsal-ventral patterning protein tolloid; astacin homology; C1r/C1s rep

Query Match 3.8%; Score 640.5; DB 2; Length 1070;
 Best Local Similarity 24.2%; Pred. No. 3.3e-28;
 Matches 222; Conservative 135; Mismatches 331; Indels 231; Gaps 33;

269 VAPGTEIR-----QSGGDP---GIPAYGRREGSRFHNGDTLKECP--AFELV 313
 Db 332 IMPGKESEFNILDEKVDLSLEPYDYSIMYSRDKFSKHYLDITRRFRQGMIALRI 391
 QY 314 GQKAITCQKNQMSAKKGCVPSC--FNFNTPSGCVLSNYPEDYG--NHLCVWLILA 369
 Db 332 GQNT---RLNDGVRQTNKLYKCPCTGRTWESGITSF---DSGIRGKTCOMRIIA 443
 QY 370 RPESRIHLAFNDIVER-QPFLVYKDGATAPLIGFSGNOLPSSITSGHVARLEFG 428
 Db 444 SHGERIQSLTRLDNSCDTIVYVRDGHFPGSLSLGFCGKKLPPEMISSTRLWVEYK 503
 QY 429 TDHSTGKRGFNITFTFRHNECPDPGVVNGKRFDSLIQSSISFLCDEGFLTGQSET 488
 Db 504 S-RAAREAFQAAF----- 516
 QY 489 ITCVLKESVYVWNSAVLRCEAPCCGHLTSPSGTILSPGMPGFYKDALSCANVIAQPGY 548
 Db 517 -----EALCGNMPPGFGFLNSPVPDEYSDKVCCEWVITVREYQ 557
 QY 549 IKITFDRKTEVN---YDLEVRDGRYSAPLIGVNGTPOVPOFLISTSVYLLFSTD 604
 Db 558 VALERATETEPDDCAVDYVIRIDGPKSLVGYTCGRTPMAISTSHLVKVPD 617
 QY 605 KSHDIDQLRYETITLQSDHCLDPGIPVNGQRHNDYF-VGALVTF--SCDSGYTL-SD 660
 Db 618 ESMQKGFPSASY--LEEVNED-----EEDHGCHVCAVNTLGSYECKIKIGYELHSD 667
 QY 661 GEPLCEBNFQMSRALPSCEALGGGFIQSSGTLISGFPDFYNNMNCWIIETSGKG 720
 Db 668 GK-----KCEACGSGYLDAPSGTISSPFDLPPDKNCVWHI--SAPKG 710
 QY 721 VFTFHTFHL-----ESGHVLLITENGSTFOPLR-QLTGSRLPAPISAGLYGNFTAQ 772
 Db 711 HTLVNFTHMDLEWRGDECELDIVRVTVNNGKNERLQGGYCGMAPPSITS--LSN-ELR 767
 QY 773 VRFISDSMSYEGNITFSEYDLPECEBEVPAVSTRKGLQFGVGDTLTSCFPGYLEG 832
 Db 768 IEFRRSDTLQKTGSMDYVA-DVDECASSNGGKHCICENT---VG-SFHCSCREGFIAD 822

QY 833 TARITCGRRRLWSSPLPRCVACGNSVTGOTLLSPNFPVNNNNHCEIYSIOTOPG 892
Db 823 DEKCKEKG-----CHYEVTDTKGVIQSPDYPSFYPARRDCEMHFTTAPG 867
QY 893 KGICLKARAFELS-----EGDVLKYDGNNSARLLGVFSHSEMMGVTLNSTSSLMWDF 947
Db 868 HVRRLIFTDQVEBHRTCRYDHVEAFDQANIQAPQIGKYGCEKAPIL-SENITLTLTF 926
QY 948 ITDAENTSKGFEHLFSSFELIKCEDPGTPKFGYKVHDEGHFAGSSVSFSCDPGYSLRGE 1007
Db 927 LSDTSVORKGF-----RARN 941
QY 1008 ELCLSGRRRTWDRPLPTCVACGSGTVRGEVSGQVLSRGPAPYEHNLNCIWTIEAEGC 1067
Db 942 DTVCOS-----PTAT-----SAPKILSHVLYGS--KPYDNRONCSWNIQAPBQ 985
QY 1068 TIGLHPLVFDTSE---VHDVLRIMDGFVESGVLLKELSGPALPKDLHSTFNSVVLQFST 1123
Db 986 HVELRFTAFEIEQGRCLYDVAVYDGPTEMDLVLGKRCGNQVPSPIVSSTRSLVFRS 1045
QY 1124 DFTSKOGFAIQFSVSTAT 1142
Db 1046 DDRIKSGGFSATYRIADDT 1064

Search completed: October 18, 2004, 10:25:19
Job time : 58 secs

This page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 18, 2004, 15:00:41 ; Search time 2762 Seconds

(without alignments)
5748.994 Million cell updates/sec

Title: US-10-016-248-2

Perfect score: 16985

Sequence: 1 MAGAPPALLPCLSLSDCC.....RSGPVGDPTSLPSHSPK 3104

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3403857 seqs, 2557783690 residues

Total number of hits satisfying chosen parameters: 6807714

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=x1h
-Q=/cgn2_1/USPro.spool/US10016248/runat_18102004_105501_11632/app_query.fasta_1.3271
-DB=Published Applications NA -QFMT=fastap -SUFFIX=tmpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=pct -FORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10016248@cgn_1.1.1858@runat_18102004_105501_11632
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOC=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	16985	100.0	10136	16	US-10-016-248-1	Sequence 1, Appl1
2	14142.5	83.3	8010	16	US-10-016-248-3	Sequence 3, Appl1
3	9077.5	53.4	6004	16	US-10-467-433-38	Sequence 38, Appl1
4	7643.5	45.0	7323	15	US-10-276-934-5	Sequence 5, Appl1
5	7643.5	45.0	8034	15	US-10-276-934-6	Sequence 6, Appl1
6	7610.5	44.8	6409	15	US-10-276-934-3	Sequence 3, Appl1
7	7159	42.1	5598	15	US-10-276-934-1	Sequence 1, Appl1
8	6991	41.2	6145	15	US-10-276-934-2	Sequence 2, Appl1
9	6978	41.1	5667	15	US-10-276-934-4	Sequence 4, Appl1
10	4148.5	24.4	4506	16	US-10-467-042-27	Sequence 27, Appl1
11	3942	23.2	2487	9	US-09-799-514-3	Sequence 3, Appl1
12	2694	15.9	2609	17	US-10-399-455-31	Sequence 69, Appl1
13	2639	15.5	2387	17	US-10-399-037-69	Sequence 20, Appl1
14	1940.5	11.4	1615	16	US-10-451-010-20	Sequence 31, Appl1
15	1852	10.9	11230	9	US-09-911-842-3	Sequence 3, Appl1
16	1852	10.9	11230	13	US-10-150-821-3	Sequence 3, Appl1
17	1797.5	10.6	10878	9	US-09-911-842-1	Sequence 1, Appl1
18	1797.5	10.6	10878	13	US-10-150-821-1	Sequence 1, Appl1
19	1796.5	10.6	11158	15	US-10-028-248A-5	Sequence 5, Appl1
20	1796.5	10.6	11158	16	US-10-028-248A-7	Sequence 5, Appl1
21	1793.5	10.6	11152	17	US-10-603-283-1	Sequence 1, Appl1
22	1792.5	10.6	11152	15	US-10-028-248A-7	Sequence 7, Appl1
23	1792.5	10.6	11152	16	US-10-107-782-7	Sequence 7, Appl1
24	1782	10.5	10674	15	US-10-295-027-429	Sequence 429, App
25	1782	10.5	10674	16	US-10-295-027-1080	Sequence 1080, App
26	1638	9.6	3810	16	US-10-108-260A-814	Sequence 814, App
27	1616	9.5	3145	16	US-10-114-270-31	Sequence 31, Appl1
28	1481	8.7	9279	13	US-10-087-192-521	Sequence 521, Appl1
29	1466.5	8.6	11920	17	US-10-741-601-51	Sequence 51, Appl1
30	1421.5	8.4	6153	9	US-09-898-570-21	Sequence 21, Appl1
31	1421.5	8.4	6153	10	US-09-839-446-21	Sequence 21, Appl1
32	1411	8.3	10872	16	US-10-152-318A-1984	Sequence 1984, App
33	1371	8.1	6951	16	US-10-641-643-1265	Sequence 1265, App
34	1371	8.1	6951	17	US-10-283-975A-293	Sequence 293, App
35	1371	8.1	7313	15	US-10-037-270-259	Sequence 259, App
36	1371	8.1	7313	15	US-10-117-722-259	Sequence 259, App
37	1370	8.1	3991	16	US-10-311-623-24	Sequence 24, Appl1
38	1366.5	8.0	7028	16	US-10-376-774-802	Sequence 802, Appl1
39	1314.5	7.7	1384	16	US-10-416-314-64	Sequence 64, Appl1
40	1270	7.5	3804	9	US-09-764-893-21	Sequence 43, Appl1
41	1270	7.5	3804	9	US-09-764-881-21	Sequence 21, Appl1
42	1270	7.5	3804	9	US-09-764-853-214	Sequence 214, Appl1
43	1270	7.5	3804	9	US-09-764-898-60	Sequence 60, Appl1
44	1270	7.5	3804	10	US-09-764-881-21	Sequence 21, Appl1
45	1270	7.5	3804	11	US-09-764-875-275	Sequence 275, Appl1

ALIGNMENTS

RESULT 1
US-10-016-248-1
Sequence 1, Application US/10016248
Publication No. US20040033491A1
GENERAL INFORMATION:
APPLICANT: Alsbjork et al.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-218
CURRENT APPLICATION NUMBER: US/10/016,248
PRIOR FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: 60/254,329
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/291,037
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: 60/255,648
PRIOR FILING DATE: 2000-12-14
PRIOR APPLICATION NUMBER: 60/297,173
PRIOR FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: 60/309,258
PRIOR FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: 60/326,393
PRIOR FILING DATE: 2001-10-01

PRIOR APPLICATION NUMBER: 60/315,639
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 167
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 10136
TYPE: DNA
ORGANISM: Homo sapiens
US-10-016-248-1

Alignment Scores:
Pred. No.: 0 Length: 10136
Score: 16985.00 Matches: 3104
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0

US-10-016-248-2 (1-3104) x US-10-016-248-1 (1-10136)

QY 1 MetAlaGlyAlaProProAlaLeuLeuLeuProCysSerLeuLeuLeuSerAspCysCys 20
DB 1 ATGGCGGGGGCCCCCTCCCGGCTTGTCTGTCTGTGCAATTGATCTGAGCTGCTGT 60
QY 21 AlAserAnglnAgnHisSerValGlyValGlyProSerGluLeuValIlyLysGlnIle 40
DB 61 GCTAGCAATCAGGACGACTCGTGGGCGCTAGAACCTCCGAGCTAGTCAGAAAGCAATT 120
QY 41 GlnLeuLysSerArgGlyValIlyLysLeuMetProSerLysAspAsnSerGlnLysThrSer 60
DB 121 GAGTTGAAGTCTGAGGTGTGAGCTGATGCCAGCAAGAACAGCCAGAGAGAGCTGT 180
QY 61 ValLeuThrGlnValGlyValSerGlnGlnIlyHisAsnMetCysProAspProGlyIlePro 80
DB 181 GTGTTAATCAGAGTGTGTGTCTCCCAAGACATATATGTGTCTCAAGCTTGGCATACC 240
QY 81 GlnArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
DB 241 GAAGAAGGGGCAAAAGACTAGAGCTCGGATTTTCAGGTTAGATCCAGCGTCACCTCG 300
QY 101 AsnGlnGlyLysThrPheLeuGlnGlySerLysArgIleThrCysMetLysValSerAspMet 120
DB 301 AACGAGGGCTATACCTCGCAAGGGTCCCAAGCGATCACCTGTATGAAGTGAAGCCGACATG 360
QY 121 PheAlaAlaThrSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisLeu 140
DB 361 TTTCGGCCCTCGAGCCACACCAAGCCAGTCTGCCGAGCCGCGATGTGTATGCCCTT 420
QY 141 ArgGlyProSerArgIlyIleIleThrSerProAsnPheProIleGlnIlyAspAsnAla 160
DB 421 CGAGGCCCCCTCGGGCATCATCCTCCCAATTTCCCATTCAGATGACAAACATATGCA 480
QY 161 HisCysValThrIleIleThrAlaLeuAsnProSerLysValIleLysLeuAlaPheGln 180
DB 481 CATCTGTGTGATCATCATCAGACATCAACCTCCCAAGGTATCAAGCTCGCTTTGAG 540
QY 181 GlnPheAspLeuGlnArgGlyLysArgPheThrLeuThrValGlyAspGlyGlnAspGly 200
DB 541 GACTTTGATTTGAGAGGGGCTATGACACCTTGACCGCTGGTCAATGTGTGTCAGATGG 600
QY 201 AspGlnLysThrValLeuIlyMetSerGlnAsnAlaCysSerAspSerProHisThrPro 220
DB 601 GACCAAAAGACAGTTCTTCAATCATGTCTCAAAATGCTCGACATGACAGCCCTCACACCCA 660
QY 221 GlySerArgIleProGluSerMetSerGlyAspIleThrArgGlnLysThrPheThrValLeu 240
DB 661 GGTCTTCGCAATCCCAAGAGCATGTCTGGGAGATCTGAGGAGCAAAATGACTGTAAT 720
QY 241 GlnIleCysArgAspIleSerSerSerAspAlaArgSerGlySerValArgLysSerPro 260
DB 721 GAGATCTGTCTGATCACTTGAAGATTGCAATGCAAGGTTCAGTTCAGTGAAGAAAGTCTCA 780
QY 261 LysThrSerAsnAlaValGlnLeuValAlaProGlyThrGlnIleGlnGlnLysSerCys 280

DB 781 AAGACTTTAATGCTGTGAACTTTGTCTCTTGGACAGAAATCAGAGGGCACTTGC 840
QY 281 GlyAspProGlyIleProAlaIleArgGlyArgArgGlyLysSerArgPheHisArgIlyAsp 300
DB 841 GGTGACCTGTGCAATACCTCATATATGCGGAGGAAAGCTCCCGGTTTACCAACGGTATC 900
QY 301 ThrLeuLysPheGlnLysGlnProAlaPheGlnLeuValGlyGlnLysAlaIleThrCys 320
DB 901 ACACTCAAGTTTGAAGTCCAGCCCGCTTTGAGCTGTGGAGACAGAAAGCAATCATATCC 960
QY 321 GlnLysAspAsnGlnThrSerAlaLysLysProGlyCysValPheSerCysPhePheAsn 340
DB 961 CAAGAAGATACCAATGCTGCGCTTAAGAACCAAGCTGCTTTCTCTGCTTTTCAAC 1020
QY 341 PheThrSerProSerGlyValValLeuSerProAsnIlyProGluAspIlyGlnHis 360
DB 1021 TTCACAGCCCGTGTGGGTTGTCTCTCCCACTACCCAGAGACTATGGCAACAC 1080
QY 361 LeuHisCysValThrLeuIleLeuAlaArgProGluSerArgIleHisLeuAlaPheAsn 380
DB 1081 CTTCAGTGTGTGCTCATCTGTGCGAGGCTGAGAGCCGATCCACCTGCTTCAAC 1140
QY 381 AspIleAspValGluProGlnPheAspPheLeuValIleLysAspGlyValaThrAlaGlu 400
DB 1141 GACATTTAGTGAAGCTCAGTTGATTTCTGTGTCAATCAAGATGGGGCCACCCCGAG 1200
QY 401 AlaProValLeuGlyThrPheSerGlyAsnGlnLeuProSerSerIleThrSerSerGly 420
DB 1201 GGGCCGCTCTGGGACCTTCTCAGAAACCAAGCTTCTCCCTCCATCAACAAGCAGTGGC 1260
QY 421 HisValAlaArgLeuGluPheGlnThrAspHisSerThrGlyLysArgIlyPheAsnIle 440
DB 1261 CACGTGGCCCTCTGAGTTCAGACTCACTCACTCAAGGAGAGAGGGGCTTCAACATC 1320
QY 441 ThrPheThrThrPheArgHisAsnGluCysProAspProGlyValProValAsnGlyLys 460
DB 1321 ACTTTTACACCTTCGACACACAGAGTCCCGATCTCGGGCTTCAGTAAATGGCAAA 1380
QY 461 ArgPheGlnLysSerLeuGlnLeuGlySerSerIleSerPheLeuCysAspGluPhe 480
DB 1381 CGGTTTGGGAGACAGCTCAGCTGGGCACTCCATCTCTCTCTGTATGAAAGCTTTC 1440
QY 481 LeuGlyThrGlnGlySerGluThrIleThrCysValLeuLysGlnGlySerValIlePro 500
DB 1441 CTGGGACTCAGGGCTTCAGAGCATCATCTGTCTCTGAAGAGAGGGAGCGGTGTGG 1500
QY 501 AsnSerAlaValLeuArgCysGlnAlaProCysGlyGlnHisLeuThrSerProSerGly 520
DB 1501 AACAGCGCTGTGCTGCGGTGAGAGCTCCGTGTGTGTCACTTGACCTTCCGCCAGCGGC 1560
QY 521 ThrIleLeuSerProGlyIlyProGlyLysThrIlyAspAlaLeuSerCysAlaIleProVal 540
DB 1561 ACCATCTCTCTCGGGCTGGCTGTCTTCAAGAGTGCCTTGAAGTGTGGCTGGGTG 1620
QY 541 IleGluAlaGlnProGlyIlyProIleLysIleThrPheAspArgPheLysThrGlnVal 560
DB 1621 ATTAGGCCCAAGCAGGCTACCATCAAAATCATCTTGCACAGATTTAAAAACGAGGTC 1680
QY 561 AsnIlyAspThrLeuGlnValArgAspGlyValArgThrIlySerAlaProLeuIleGlyVal 580
DB 1681 AACTATGACACCTGGAAGTACGCGATGGGAGACTTACTAGGCGCTTATGGGGTT 1740
QY 581 TyrHisArgIlyThrGlnValAlaProGlnPheLeuIleSerThrAsnIlyLeuIlyLeuLeu 600
DB 1741 TACACGGGAGCCAGTTTCCCAATCTCTATCAAGCACCAACCAATCACTTACTCTCTC 1800
QY 601 PheSerThrAspLysSerHisSerAspIleGlyPheGlnLeuValArgIlyGlnThrIleThr 620
DB 1801 TTCTCTACCGAAGAGTCACTGGACATGGGCTTCCAGCTCCCTATGACATTAACA 1860
QY 621 LeuGlnSerAspHisCysLeuAspProGlyIleProValAsnGlnIlyGlnArgHisGlyAsn 640

Db 1861 CTGCAGTCAGACCACTGCTGGATCCAGAAATCCCAATAAGCAGCGTCATGGGAAAT 1920
Qy AspPheTyrValGlyAlaLeuValThrPheSerCyAspSerGlyTyrThrLeuSerAsp 660
Db 1921 GACTTCAACGGGGCGCGCTGGTAGCTTCAGCTGTACCTCGGGCTACACATTAAGGAC 1980
Qy 661 GlyValProLeuGluCyAspGluProAsnPheGlnTyrPheSerArgAlaLeuProSerCyGlu 680
Db 1981 GGGAGGCTCTGAGAGTGTAGACCCAACTTCAGTGGAGCCGGGGCTGCCAGTGTGAA 2040
Qy 681 AlaLeuCyGlyGlyPheIleGlnGlySerSerGlyThrIleLeuSerProGlyPhePro 700
Db 2041 GCTCTGTGGTGGCTTCATTCAGAGGCTCCAGTGGAGCAATCTTGCGCAGGGTCCCT 2100
Qy 701 AsnPheTyrProAsnAsnLeuAsnCySerThrTyrIleIleGlnTyrSerHisGlyValGly 720
Db 2101 GACTTCAACCCCAACCACTGAACTGACCTGGAATTCGAAACATCTCAATGGCAAGGT 2160
Qy 721 ValPhePheThrPheHisThrPheHisLeuGluSerGlyHisAspTyrLeuLeuIleThr 740
Db 2161 GTGCTTTCATCTTCACACCTTCACCTGGAAAGTGGCCATGACTACCTCTCATACT 2220
Qy 741 GluAsnGlySerPheThrGlnProLeuArgGlnLeuThrGlySerArgLeuProAlaPro 760
Db 2221 GAGAACGGCAGCTTCACCCAGCCCTGAGGAGCTAACTGGATCTCGGCTGCCAGTCCC 2280
Qy 761 IleSerArgLeuTyrGlyAsnPheThrAlaGlnValArgPheIleSerAspPheSer 780
Db 2281 ATCAGCCCTGGGCTCTATGGCACTTCAGCCAGGTCGCGCTTCATCTGATTTCTCC 2340
Qy 781 MetSerTyrGluGlyPheAsnIleThrPheSerGluTyrAspLeuGluProCyGlyGlu 800
Db 2341 ATGCAATATGAAGATTCACACATCACTTCTCAGAGTACGACTTGGAGCCCTGGAGAG 2400
Qy 801 ProGluValProAlaTyrSerIleArgValGlyLeuGlnPheGlyValGlyAspThrLeu 820
Db 2401 CCCGAGGCTCCAGCCTACAGCATCCGGAAGGCTTGCACTTGCGTGGGAGCACCTTG 2460
Qy 821 ThrPheSerCyAspPheProGlyTyrArgLeuGlnGlyThrAlaArgIleThrCyAspLeu 840
Db 2461 ACCCTTCCTGCTTCCCGGGGTACCGTGGAGGGCACCGCCCATCAGCTGCTGGGG 2520
Qy 841 GlyArgArgArgLeuTyrPheSerProLeuProArgCyValAlaGluCyGlyAsnSer 860
Db 2521 GCGCAGAGCGGCGCTGTGAGACTCCCTCTGCCAAGGTGTGTGCTGAGTGGGAATTCA 2580
Qy 861 ValThrGlyThrGlnGlyThrLeuLeuSerProAsnPheProValAsnTyrAsnAsnAsn 880
Db 2581 GTCAACAGCACTCAGGGGACTTGTGCTGCCCAACTTCTCTGGAATCAATCAAT 2640
Qy 881 HisGluCysIleTyrSerIleGlnThrGlnProGlyValGlyIleGlnLeuValAlaArg 900
Db 2641 CATGAATGCACTCTACTCCATCCAGACCAGCAGGAAAGGAAATTCAGCTGAAAGCCAGG 2700
Qy 901 AlaPheGluLeuSerGlyValAspValLeuValTyrAspGlyAsnAsnAsnSerAla 920
Db 2701 GCATTCGAATCTCTCGAAGAGATGCTCTCAAGGTTATATGAGCAACCAACTCCGCC 2760
Qy 921 ArgLeuLeuGlyValPheSerHisSerGluMetMetGlyValThrLeuAsnSerThrSer 940
Db 2761 CGTTTGTGGGAGTTTTTAAGCATTTGAGATGATGGGGGTGACTTTGAAACAGCAATCC 2820
Qy 941 SerSerLeuTyrLeuAspPheIleThrAspAlaGluAsnThrSerGlyPheGlnLeu 960
Db 2821 AGCAGTGTGGCTTGATTTTCATCATCATGAGTCTGAAACACCAAGCAAGGGCTTTGAACTG 2880
Qy 961 HisPheSerSerPheGluLeuIleValGlyCyGlyLeuAspProGlyTyrProLysPheGlyTyr 980
Db 2881 CACTTTTCCAGCTTTGAATCATATCAAAATGTAGAGACCCAGAAACCCCAAGTTTGGCTAC 2940
Qy 981 LysValHisAspGluGlyHisPheAlaGlySerSerValSerPheSerCyAspProGly 1000
Db 2941 AAGTTTATGATGAAGGTCAATTTTGACAGGAGGCTCCGTTCTTCACTGTGACCTTGGA 3000

Qy 1001 TyrSerLeuArgGlySerGluGluLeuLeuCysLeuSerGlyGluValArgTyrTrpAsp 1020
Db 3001 TACAGCTCGGGGAGTGTAGAGAGCTGCTGTGTCTAGTGTGAGAGCGCCGACTGGAG 3060
Qy 1021 ArgProLeuProThrCysValAlaGluCyGlyGlyThrValArgGlyGluValSerGly 1040
Db 3061 CGGCTCTGCGCACTGTGTCCGAGGTGTGAGGAGACGTAGAGAGAGAGGTGTGGGG 3120
Qy 1041 GlnValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysIleTyrThr 1060
Db 3121 CAGGTCTGTACCCCGGATTCAGCTCCCTAAGAACACAAATCTCAACTGCATCTGAGCC 3180
Qy 1061 IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu 1080
Db 3181 ATGAAACAGAGGCGGCTGACCATTTGGAGTCACTTCTGTGTGTGTGACACAGAGAG 3240
Qy 1081 ValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyValLeuLeuGlyGlu 1100
Db 3241 GTTCAAGAGGTGCTGGCATCTGGATGGGCTGTGGAGAGCGGGTCTGTGAAGAGAG 3300
Qy 1101 LeuSerGlyProAlaLeuProLysAspLeuHisSerThrPheAsnSerValValLeuGln 1120
Db 3301 CTGAGTGGCCCGGCTCTGCCAAGACCTGCATAGACCTTCAACTCGTCTGTCAG 3360
Qy 1121 PheSerThrAspPheThrSerLeuGlyGlnGlyPheAlaIleGlnPheSerValSerThr 1140
Db 3361 TTCAAGCATGACTTCTTACACAGACAGAGGCTTTGGCATTAATTTTCAGTGTCCACA 3420
Qy 1141 AlaThrSerCyAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr 1160
Db 3421 GCAACGTCCTGCATATGACCTGGGATCCGCAAAATGGAGTGGAGTGTGACAGTTGG 3480
Qy 1161 GluAlaGlyAspSerThrValPheGlnCyAspProGlyTyrAlaLeuGlnGlySerAla 1180
Db 3481 GAAGCGGCGCACTCCACAGTGTTCAGTGTGACCTGGCTGACCGCTGCGAGGAAAGTCA 3540
Qy 1181 GluIleSerCyValIleValIleGluAsnArgPheThrGlnProSerProProThrCys 1200
Db 3541 GAGATGAGCTGTGTGAAGATCGAAACAGGTTTCTTGGAGGCCACCGCCCAACATGC 3600
Qy 1201 IleAlaProCyGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr 1220
Db 3601 ATGCTTCCCTGGGGGAGACCTGCACAGACCATCTGAGTATCTCTCACCMAATTAC 3660
Qy 1221 ProGluProTyrProProGlyValGlyCyAspTyrLysValThrValSerProAspTyr 1240
Db 3661 CCAGAAACCTTACCCGCGCAGGCAAGAGTGTGACTGGAAGTGAACGCTTCCACAGACTAC 3720
Qy 1241 ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle 1260
Db 3721 GTCATGGCCCTGGTATTTTAACATCTTAACTGTGAGCCGTGGCATATATCTTCCCAATAC 3780
Qy 1261 TyrAspGlyArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro 1280
Db 3781 TAGCAGGAGGGAATCTCTCAGCCCTCTCATATGAAAGCTTATAGCTCCAGCTCCA 3840
Qy 1281 GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal 1300
Db 3841 GCGCGCATTTGAAGACGACAGCAACAGCTTCTCTGCGCTTCCGACGATGATCTGTG 3900
Qy 1301 SerAspAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCyAspPheAspPro 1320
Db 3901 AGCAATGTGGCTTGTGATGCTATGACGAAACCCGGGGAGTCAATGTTTGAATCT 3960
Qy 1321 GlySerIleValAsnGlyThrArgValGlySerAspLeuValLeuGlySerSerValThr 1340
Db 3961 GTTTCATCAAGAACGAGCACCGGTTGGGCTCGACTGAAGTGGGCTCTCTCGTACAC 4020
Qy 1341 TyrTyrCysHisGlyGlyTyrGluValGluGlyThrSerThrLeuSerCysIleLeuGly 1360
Db 4021 TACTACTGCAACGGGGCTTCGAAAGTTGAGGCACTCGACCTGAGCTGCATCTGGGG 4080

Qy	1361	ProaapgllylserprovaltrpaaasnproarprovalcysThrAlaProCyaglly	1380
Db	4081	CCTGATGGGAAGCCCGTGTGAAACAATCCCGGCCAGTCTGCACAGCCCTGTGGGGA	4140
Qy	1381	GlnTyValIglYserAspGlyValIleuSerProaenTyProGlnAsenTyThrSer	1400
Db	4141	CAGTATGGGGTTCGAGCGGAGTGTCTTGTCCCACTACCCCACTAACCACTACACAGT	4200
Qy	1401	GlyGlnIleCybleuTyPheValThrValProIysAspTyValIlePheGlyGlnPhe	1420
Db	4201	GGACAGATCTGCTGTATTTGTATCTGTGCCAAGACATAATGTTGGTGGCAATTC	4260
Qy	1421	AlaPhePheHisThrAlaLeuAsnAspValGluValHisAspGlyHisSerGlnHis	1440
Db	4261	GCTTCTTTCAACAGGCCCTCAACGAGTGTGAGGTTCACGACGGCCACAGCCAGCAC	4320
Qy	1441	SerArgLeuLeuSerSerLeuSerGlySerHisThrGlyGluSerLeuProLeuAlaThr	1460
Db	4321	TGGGGGCTCCCTCAGCTCCCTCGGGCTCCCATACAGAGATCATGCTGGCCAC	4380
Qy	1461	SerAsnGlnValIleuIleLeuPheSerAlaLeuGlyLeuAlaProIleAspGlyPheHis	1480
Db	4381	TCCATACAGATTCATTAAGTATGAGCCCAAGGCCCTCGACACGACAGAGGCTTCAC	4440
Qy	1481	PheValTyrglnAlaValProArgThrSerAlaThrGlnCyserSerValProGluPro	1500
Db	4441	TTTGTCTACCAAGCGGTTCTCGAACGAGGCCACGAGTGAGCTGTGTCCGGAAACC	4500
Qy	1501	ArgTyrglyAsnGluGlySerAspPheSerValGluValAlleValArgPheGlyCys	1520
Db	4501	CGCTATGCAAGAGCGCTGGGAGTACTTCTGGTGGGGGCCATCGTCCCTTCGATGC	4560
Qy	1521	AsnSerGlyTyralaLeuGlnGlySerProGluIleGluCysLeuProValProGlyAla	1540
Db	4561	AACCTCGGCTATGCTCCGACAGGGGTCCCGACAGATCAGAGCTCTCGTGTGGAGCC	4620
Qy	1541	LeuAlaGlnTrpAsnValSerAlaProThrCysValValProCyagllyAsnLeuThr	1560
Db	4621	TTGGCCCAATGAGATGTCTCAGCGCCCACTGTGTGGTCCGTTGGAGCAACTGCACA	4680
Qy	1561	GluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrlleuAsnSerIleuAsn	1580
Db	4681	GAGGCGAGGGGACCACTCTGTCCCTGGCTCCCGAGCGGATCTCAACACCTCAAC	4740
Qy	1581	CysValTrpIysIleValIleProGluGlyAlaGlyIleGlnIleGlnValIleSerPhe	1600
Db	4741	TGTGTGGAAAGATCGTGTCCCGAAAGCGCTGGATCCAGATCCAAATGTGCACTTT	4800
Qy	1601	ValThrGlnGlnAsnTrpAspSerLeuGluValPheAspGlyAlaAspAsnThrValThr	1620
Db	4801	GTCACAGACGAACTGGAGCTCGCTGGAATTTGATGTGTCAGATTAACACTGTAACT	4860
Qy	1621	MetLeuGlySerPheSerGlyThrThrValProIleLeuAsnSerThrSerAsnGln	1640
Db	4861	ATCTGGGAGGTTTCTCAGAAACAACCGTCTGCTCTTGTGAACGCACTTCAACACG	4920
Qy	1641	LeuTyrlleuHisPheTyrlleuSerAspIleSerValSerAlaAlaGlyPheHisIleuGlnTyrl	1660
Db	4921	CTCTACCTTCAATTTCTACTCAGATATCAGGTATCTGCAAGCTGGCTTCACTTGAAGTAC	4980
Qy	1661	IysThrValGlyIleuSerSerCysProGluProAlaValProSerAsnGlyValIysThr	1680
Db	4981	AAAACGATGGGCTGTGAGAGTGTTCGGAACCTGCTGCCCACTGAACGGGGTGAAGCT	5040
Qy	1681	GlyGluArgTyrlleuValAsnAspValIleSerPheGlnCysGluProGlyTyrlAlaLeu	1700
Db	5041	GGGAGACGCTACTTGTGTAATGATGTGTCTTTTCCAGTGTGAGCCGGGAATATGCCCTC	5100
Qy	1701	GlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTrpAsnTyrlProPro	1720
Db	5101	CAGGGCCAGCCCACTCTCTGCAATGCCGGAACAGTGGCGATGGAACATACCTCTCT	5160
Qy	1721	ProLeuCyrlleuIleGlnCysGlyGlyThrValGluGluMetGlnGlyValIleLeuSer	1740

Db	5161	CACTCTGTATTTGACAGTGTGGGGAACAGTGGAGGAATGAGAGGGGTGATCTCGAC	5220
Qy	1741	ProGlyPheProGlyAsnTyrlProSerAsnMetAspCysSerTrpIysIleAlaLeuPro	1760
Db	5221	CCGGCTTCCAGGGAACATACCCCAAGTAACTGAGTGTCTCTGAAAAATAGCATCTGCC	5280
Qy	1761	ValGlyPheGlyAlaHisIleGlnPheLeuAsnPheSerThrGluProAsnHisAspTyrl	1780
Db	5281	GTTGGCTTTGAGCTCACATTCAGTTCCTGAACTTCCACAGGACCCCAACACAGCTAC	5340
Qy	1781	IleGluIleArgAsnGlyProTyrglyIlePheSerArgMetMetGlyArgPheSerGlySer	1800
Db	5341	ATAGAAATCCGAATGGCCCTTATAGAGACACGCCCATGATGGAAGATTCAGTGAAGC	5400
Qy	1801	GluLeuProSerSerLeuLeuSerThrSerHisGlnThrThrValTyrlPheHisSerAsp	1820
Db	5401	GAGCTTCCAGCTCTCTCTCTCCACGTCCACGAGACCAACCGTGTATTTCCACAGCCAC	5460
Qy	1821	HisSerGlnAsnArgProGlyPheIysIleuGlyTyrglnAlaTyrglyIleuGlnGlyCys	1840
Db	5461	CACTCCAGAAATCGGCCAGGATTCAGCTGGAGTATCAGGCTTATGAACCTTCAAGAGTGC	5520
Qy	1841	ProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTyrlAsnValGlyGln	1860
Db	5521	CCAGACCCAGAGCCCTTTGCCAATGCACTTGTAGGGAGGCTGGCTACACGTGGACAA	5580
Qy	1861	SerValThrPheGluCysLeuProGlyTyrglyIleuThrGlyHisAspValIleuThrCys	1880
Db	5581	TCAGTACCTTCGAGTGTCTCCGGGGTATCAATGATGGCCACCTGTCTTCACTGCT	5640
Qy	1881	GlnHisGlyThrAsnArgAsnTrpAspHisAspLeuProIysCysGluValProCyaglly	1900
Db	5641	CAACATGGACCAACCGGAACCTGGAGCACCCCTGCCAAGTGAAGTCCCTGTGGC	5700
Qy	1901	GlyAsnIleThrSerSerAsnGlyThrValTyrlSerProGlyPheProSerProTyrlSer	1920
Db	5701	GGGAACATCACTTCTTCAAGGCACTGTATCTCCCGGGGTTCCTTAGCCGTACTCC	5760
Qy	1921	SerSerGlnAspCysValTrpLeuIleThrValProIleGlyHisGlyValArgLeuAsn	1940
Db	5761	AGCTCCAGAGCTGTGTCTGTGATGATCACTGTGCCCATTTGGCGCTGGCTCAAC	5820
Qy	1941	LeuSerLeuLeuGlnIleThrGluProSerGlyAspPheIleThrIleTrpAspGlyProGln	1960
Db	5821	CTCAGCTGTGCAACAAGACCTCTGGAATTTCAACATCTGGAGTGGGCCACAG	5880
Qy	1961	GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaIysIysThrValGlnSer	1980
Db	5881	CAACAGACCAACGCGCTCGGCTCTTCAACCGGAGCATGGCCAGAAACAGTGCAGAGT	5940
Qy	1981	SerSerAsnGlnValIleLeuIysPheHisArgAspAlaAlaThrGlyGlyIlePheAla	2000
Db	5941	TCATCCAAACGAGTCTGCTCAAGTTCACCGTATCAGCAACGAGGGGAGATCTTGCC	6000
Qy	2001	IleAlaPheSerAlaTyrlProLeuThrIysCysProProProThrIleLeuProAsnAla	2020
Db	6001	ATACCTTCTCGCTTATTCATCTACCAATGCCCTCTCCACATCTCCCAACAGCC	6060
Qy	2021	GluValValThrGluAsnGluGluPheAsnIleGlyAspIleValArgTyrlArgCysLeu	2040
Db	6061	GAAGCTGTACAGAGATGAAGATTCATATATGATGATCAATCGAGCTACAGATGCTC	6120
Qy	2041	ProGlyPheThrIleuValGlyAsnGluIleLeuThrCysIysIleuGlyThrTyrlLeuGln	2060
Db	6121	CTGGCTTTTACCTTAAGTGGGAATGAATTTCTACCTGCAAACTTGGAACTTCTGAC	6180
Qy	2061	PheGluGlyProProPheIleCysGluValHisCysAspThrAsnGluLeuLeuThrAsp	2080
Db	6181	TTTGAAGACACCCCGGATATGTGAATGCACTGTCCAAACAAATGAGCTTCTGACAGAC	6240
Qy	2081	SerThrGlyValIleLeuSerGlnSerTyrlProGlySerTyrlProGlnPheGlnThrCys	2100

Db	6241	TCACAGGCGTGAATCTGAGCCAGAGCTACCCCTGGAAAGCTATCCCAAGTTCACAGCTGC	6300
Qy	2101	SetrPleuValArgValAGluProAspPyrAsnIleSerLeuThrValGluTyrPheLeu	2120
Db	6301	TCCTGGCTGGAGAGTGAGACCCGACTATACATCTCCACAGAGAGATCTCTCTC	6360
Qy	2121	SerGluYsgInTyrAspGluPheGluIlePheAspGlyProSerGlyGlnSerProLeu	2140
Db	6361	AGCGAAGACCAATATGATGAGTTTGATGATTTTATGATGATTCATGAGACAGAGTCTCTG	6420
Qy	2141	LeuIysAlaLeuSerGlyAsnTyrSerAlaProLeuIleValThrSerSerAsnSer	2160
Db	6421	CTGAAGGCCCTCAGTGGGAATTAATCACTCCCGATTTGTCACAGCTCAAGCAACTCT	6480
Qy	2161	ValTyrLeuArgTyrPseSerAspHisAlaTyrAsnAGlyGlyPheIysIleArgTyr	2180
Db	6481	GTGATACCTGGCTGGTCACTGATCAACCCCAATCGAAGGGCTTCMAATCCGCTAT	6540
Qy	2181	SerAlaProTyrCysSerLeuProArgAlaProLeuHisGlyPheIleLeuGlyGlnThr	2200
Db	6541	TCAGCCCTTACTGACAGCTGCGCCAGGCTCCACTCCATGGCTTCTATGGCCAGACCC	6600
Qy	2201	SerThrGlnProGlyYsgIleHisPheGlyCysAsnAlaGlyTyrArgLeuValGly	2220
Db	6601	AGCACCCAGCCCGGGGCTCCATCCATTTGGCTGCAACGCCGGCTACCGCTGTGGGA	6660
Qy	2221	HisSerMetAlaIleCysThrArgHisArgProGlnGlyTyrHisLeuTyrPseSerGluAlaIle	2240
Db	6661	CACAGCATGGCCATCTGTACCCGCGACCCCAAGGCTACCACTGTGGAGGAAGCCATC	6720
Qy	2241	ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProIysAsnGlyMetValPhe	2260
Db	6721	CCTCTCTGTCAAGCTCTTTCTGTGGGCTTCTCTAGGCCCCCAAGATGAGATGTGTTT	6780
Qy	2261	GlyYsgIuTyrThrValGlyThrIysAlaValTyrSerCysSerGlyGlyTyrHisIleu	2280
Db	6781	GGCAAGAGGTACACAGTGGGAACCAAGCCGTGTACAGCTCAGTGAAGGCTACCACTC	6840
Qy	2281	GlnAlaGlyAlaGluAlaThrAlaGluCysAlaLeuAspThrGlyLeuTyrPseSerAsnArgAsn	2300
Db	6841	CAGCAGGCGGTGAGGCCACAGAGTGTGACACAGGCTTATGAGAGCAACCGCAT	6900
Qy	2301	ValProProGlnCysValProValThrCysProAspValSerSerIleSerValGlnHis	2320
Db	6901	GTCCACCAACAGTGTCTCTGTAATCTTCTGATGTCAATGATCAGGCTGAGAT	6960
Qy	2321	GlyArgTyrArgLeuIlePheGluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIle	2340
Db	6961	GGCGGATGGAGGCTTATCTTTGAGACACAGTATCAGTTCCAGGCCAGCTCATCTCATC	7020
Qy	2341	CysAspProGlyTyrTyrTyrThrGlyGlnArgValIleArgCysGlnAlaAsnGlyIys	2360
Db	7021	TGTACCTCGCTACTACTATACGCGCAAGGGGATCCGCTGACGGCCAAATGGCAAA	7080
Qy	2361	TrpSerLeuGlyAspSerThrProThrCysArgIleIleSerCysGlyGluLeuProIle	2380
Db	7081	TGAGACCTCGGAGACTTACGCCCACTGCGGAATCATCTCTGGAGAACCTCCGAT	7140
Qy	2381	ProProAsnGlyHisArgIleGlyThrLeuSerValTyrGlyAlaThrAlaIlePheSer	2400
Db	7141	CCCCCAATGGCCACCGCATCGGAACATGTCTGTCTACGGGGCAACAGCATTTCTCC	7200
Qy	2401	CysAsnSerGlyTyrThrLeuValGlySerArgValArgGluCysMetAlaAsnGlyLeu	2420
Db	7201	TGCAATTCGGATACACACTGTGGGCTCCAGGGGCTGTGATCATGGCCAAATGGGCTC	7260
Qy	2421	TrpSerGlySerGluValArgCysLeuAlaGlyHisCysGlyTyrProGluProIleVal	2440
Db	7261	TGAGATGCTCTGATGCTCGCTGCTTGTGTGACATGTGGGAACTCTGAGCCCAATGTC	7320
Qy	2441	AsnGlyHisIleAsnGlyGluAsnTyrSerTyrArgIysSerValIvalTyrGlnCysAsn	2460
Db	7321	AACGAGACATCATGGGGAGAACTACAGCTACCGGGGCAGTGTGTGTACATGCAAT	7380
Qy	2461	AlaGlyPheArgLeuIleGlyMetSerValArgIleCysGlnGlnAspHisHisTyrPse	2480
Db	7381	GCTGGCTTCGGCTGATGGGATGTCTGTGGCAATCTGGCAAGACAGATCATCATGTGGTCG	7440
Qy	2481	GlyYsgThrProPheCysValProIleThrCysGlyHisAspProIysAsnProValAsnGly	2500
Db	7441	GGCAAGACCCCTTCTGTGTGCCAATTAATCTGTGGACACCAAGCAACCCGTCAACGGCC	7500
Qy	2501	LeuThrGlnGlyAsnGlnPheAsnLeuAsnAspValIalIysPheValCysAsnProGly	2520
Db	7501	CTCAGTCAAGGATACCAAGTTTAACTCAAGATGTGTGCAAGTTTGTTCACACCTCGGG	7560
Qy	2521	TyrMetAlaGluGlyAlaAlaArgSerGlyCysLeuAlaSerGlyGlnTyrPseSerAspMet	2540
Db	7561	TATATGGCTGAGGGGCTGTAGGTCCCAATGTCTGCGCAGCGGGCAATGAGTACATG	7620
Qy	2541	LeuProThrCysArgIleIleAsnCysThrAspProGlyHisGlnGlnAsnSerValArg	2560
Db	7621	CTGCCCACTGGCAAGATCATCACTGTACAGATCTGTGACACCAAGAAATAGTTCGT	7680
Qy	2561	GlnValHisAlaSerGlyProHisArgPheSerPheGlyThrValSerTyrArgCys	2580
Db	7681	CAGGTCCAGCCAGCGGCCGACAGGTTCAAGCTTCGGACACACTGTGTCTTACCGGTGC	7740
Qy	2581	AsnHisGlyPheTyrLeuLeuGlyThrProValLeuSerCysGlnGlyAspGlyThrTyr	2600
Db	7741	AACCAAGCTTCTTACCTCTGGGACCCAGTGTCAAGCTGACGCGCAAGAGATGGACATGG	7800
Qy	2601	AspArgProArgProGlnCysLeuLeuValSerCysGlyHisAspProIysSerProHis	2620
Db	7801	GACGTTCGCCGCCAGCTGTCTTGTGTCTGTGTCCGTGTGATCCGGCTCCCGCTCAC	7860
Qy	2621	SerGlnMetSerGlyAspSerTyrThrValGlyAlaValIalArgTyrSerCysIleGly	2640
Db	7861	TCCAGATGTCTGGAGACATTTATCTGTGGAGAGTGTGTGGGTACACTGATGAGGC	7920
Qy	2641	LysArgThrLeuValGlyAsnSerThrArgMetCysGlyIysLeuAspGlyHisTyrPthGly	2660
Db	7921	AAGGTACTGTGTGGGAAACAGACCCGATGTGTGGCTGTGATGGACACTGGACTGGC	7980
Qy	2661	SerLeuProHisCysSerGlyThrSerValGlyValCysGlyAspProGlyIleProAla	2680
Db	7981	TCCCTCCCTCACTGCTCAGAAACCAAGCTGGAGTTGGCGTACCTGGGATCCCGCT	8040
Qy	2681	HisGlyIleArgLeuGlyAspSerPheAspProGlyThrValIleArgPheSerCysGlu	2700
Db	8041	CATGGCATCCGTTTGGGGACAGCTTTGATCCAGGCACTGTGATGCGCTTCAAGCTGAA	8100
Qy	2701	AlaGlyHisValLeuArgGlyIysSerSerGluArgThrCysGlnAlaAsnGlyIysTyrPse	2720
Db	8101	GCTGGCCACGTGTCTCCGGGATGTCTCAGACCGCACTGTGAACCAATGGCTGTGGAGC	8160
Qy	2721	GlySerGlnProGluCysGlyValIleSerCysGlyIysAspProGlyThrProSerAsnAla	2740
Db	8161	GGCTCGAGCTGAGTGTGAGTGTCTTGTGGGAACCTGGGACTCCAAGTATGCC	8220
Qy	2741	ArgValAlaPheSerAspGlyLeuValPheSerSerIleValTyrGluCysArgGlu	2760
Db	8221	CGAGTTGTGTTAGATGAGCTGTGTTTCTCCAGCTCTATCGTATGATGATGCGGGAA	8280
Qy	2761	GlyTyrTyrAlaThrGlyLeuLeuSerArgHisCysSerValAsnGlyIysTyrPthGly	2780
Db	8281	GGATACTACGCCCAAGCGCTGTCAAGCGTCACTGCTGGTCAATGGTACTGGACAGGC	8340
Qy	2781	SerAspProGluCysLeuValIleAsnCysGlyIysAspProGlyIleProAlaAsnGlyLeu	2800
Db	8341	AGTACCTTAAGTGTCTGTATTAACCTGTGTGACCTTGGATTCAGCCAAATGGCTT	8400
Qy	2801	ArgLeuGlyIysAsnAspPheArgTyrAsnTyrThrValThrTyrGlnCysValProGlyTyr	2820
Db	8401	CGGCTGGCAATGACTTCAAGTACAAACAAATGTGTACATATGATGTGTCCCTGGCTAT	8460

```

QY 2821 MetMetGluSerHisArgValSerValLeuSerCysThrLysAspArgThrTrpAsnGly 2840
DB 8461 ATGATGAGAGTCACTAGAGTATCTGTCTGAGCTGACCAAGACCGGACATGGAAATGGA 8520
QY 2841 ThrLysProValCysValAlaLeuMetCysLysProProLeuLysProAsnGlyLys 2860
DB 8521 ACCAGACCCGCTGCAAGAGCTCTCATGTGCAAGCACTCCGCTCATCCCAATGGGAG 8580
QY 2861 ValValGlySerAspPheMetTrpGlySerSerValThrValCysLeuGlnGlyTyr 2880
DB 8581 GTGTGGGCTGCTGACTTATATGAGGGCTCAAGTGTGACTTATGCTCTGAGAGGGGTAC 8640
QY 2881 GlnLeuSerLeuProAlaValPheThrCysGlnGlyAsnGlySerTrpThrGlnGluLeu 2900
DB 8641 CAGCTCTCTCCGCGCGGCTTTCACCTGTGAGGAAATGGTCTCTGACCGGAGAGACTG 8700
QY 2901 ProGlnCysPheProValPheCysGlyAspProGlyValProSerArgLysArgArgGlu 2920
DB 8701 CCTCAGTGTCTTCCCTGTCTTCTGCGGGGATCTGTGTCCGCTCCGCTGGAGAGAGAG 8760
QY 2921 AsparGlyPheSerTyrArgSerSerValSerPheSerCysHisProProLeuValLeu 2940
DB 8761 GACCGAGGCTTCTCTTACAGGTCACTGTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTG 8820
QY 2941 ValGlySerProArgArgPheCysGlnSerAspGlyThrTrpSerGlyThrGlnProSer 2960
DB 8821 GTGGGCTCTCCACCAAGGTTTGGCAGTGTGAGCATGGAGTGGCACCCAGCCAGCCAGC 8880
QY 2961 CysLeuLeuProThrLeuThrThrCysAlaAspProGlyValProGlnPheGlyTyrLeu 2980
DB 8881 TGCATTAATCCGACCCGACCACTGTCGCGGACCTGTGTGCGCACAGTTTGGAAATACAG 8940
QY 2981 AsnAsnSerGlnGlyTyrGlnValGlySerThrValLeuPheArgCysGlnLysGlyTyr 3000
DB 8941 AACATTCTCAGGCTTACAGAGTGTGAGACAGACTCTCTTCCGTTCTCAAAAAGGCTAC 9000
QY 3001 LeuLeuGlnGlySerThrThrArgThrCysLeuProAsnLeuThrTrpSerGlyThrPro 3020
DB 9001 CTGCTTTCAGGCTCTCCACCAAGGACTGCTCCCAAACTGACCTGAGTGGAAACCCCA 9060
QY 3021 ProAspCysValProHisHisCysArgGlnProGlnThrProThrHisAlaAsnValGly 3040
DB 9061 CCGACTGTGTCTCCCAACCACTGACGAGCAGCCAGCAAGCATGCGCAACGTCGCGG 9120
QY 3041 AlaLeuAspLeuProSerMetGlyTyrThrLeuLysThrProAlaArgArgLysSerPro 3060
DB 9121 GCCCTGATTTGGCCCTCATAGGGCTACAGCTCATTACTCTGCGAGAGGGCTTCTCC 9180
QY 3061 SerArgValAlaProSerThrAlaProAlaArgArgMetAlaAlaGlyGlnAlaSerArg 3080
DB 9181 TCAGAGGTGGCTCGAGACACCGACCTGCAAGGCGGATGGACACTGACAGGCAAGCCGC 9240
QY 3081 ProSerAlaTrpArgSerGlyProValGlyAspProSerThrLeuProGlySerHisArg 3100
DB 9241 CCACTGTGCTGAGAGTCCGCGCCCAATGGAGAGACCACTCAACACTGCGCGGAGCCAC 9300
QY 3101 SerProLysPro 3124
DB 9301 TCACCCAAAGCCT 9312

```

RESULT 2

```

US-10-016-248-3
; Sequence 3, Application US/10016248
; Publication No. US20040033491A1
; GENERAL INFORMATION:
; APPLICANT: Aleobrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-218
; CURRENT APPLICATION NUMBER: US/10/016,248
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/254,329
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/291,037

```

```

; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/255,648
; PRIOR FILING DATE: 2000-12-14
; PRIOR APPLICATION NUMBER: 60/297,173
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/309,258
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/326,393
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/315,639
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 8010
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-016-248-3

```

Alignment Scores:

```

Pred. No.: 0 Length: 8010
Score: 14142.50 Matches: 2605
Percent Similarity: 97.71% Conservative: 1
Best Local Similarity: 97.68% Mismatches: 2
Query Match: 83.26% Indels: 59
DB: 16 Gaps: 3

```

US-10-016-248-2 (1-3104) x US-10-016-248-3 (1-8010)

```

QY 1 MetAlaGlyAlaProProProAlaLeuLeuLeuProCysSerLeuLysSerAspCys 20
DB 1 ATGGGGGGGGGCCCCCTCCCCCTTGTCTGCTTGTGAGTTTATCTCAGACTGCTGT 60
QY 21 AlaSerAsnGlnArgHisSerValGlyValGlyProSerGlnLeuValLysGlnLe 40
DB 61 GCTACCAATCAGCGACACTCCGTGGCGGTAGACCTCCGAGCTGATCAAGAAACAATT 120
QY 41 GlnLeuLysSerArgGlyValLysLeuMetProSerLysAspAsnSerGlnLysThrSer 60
DB 121 GAGTTTAAAGTCTCGAGGTGTAAGTATGCCCCAGCAAGCAAGCAAGCAAGCAAGCT 180
QY 61 ValLeuThrGlnValGlyValSerGlnGlyHisAsnMetCysProAspProGlyTyrPro 80
DB 181 GTTTAACTCAGGTGGTGTGTCTCCAGAGACATAATATGTCTCAGACCTGGCATACC 240
QY 81 GlnArgGlyLysArgLeuGlySerAspPheArgLeuGlySerSerValGlnPheThrCys 100
DB 241 GAAAGGGGCAAAAGACTAGGCTCGAATTTCAAGTTAGGATCCAGCTCCAGTTCACTGC 300
QY 101 AsnGlnGlyTyrAspLeuGlnGlySerLysArgLysThrCysMetLysValSerAspMet 120
DB 301 AACAGGGCTATGACTTCAGAGGCTCCAGGCGATACCTTATGAAAGTGAAGGACATG 360
QY 121 PheAlaAlaTrpSerAspHisArgProValCysArgAlaArgMetCysAspAlaHisLeu 140
DB 361 TTTGGGCGCTGAGAGACCAAGGCCAGTGTGCCGAGCGGCAGATGTGATGCCACTT 420
QY 141 ArgGlyProSerGlyLysLeuThrSerProAsnPheProLysGlnTyrAspAsnAla 160
DB 421 CGAGGCCCTCGGCAATCATCACTCCCAATTTCCCAATTCAGTATGACMAAATGCA 480
QY 161 HisCysValTrpLeuLeuThrAlaLeuAsnProSerLysValLysLeuAlaPheGln 180
DB 481 CACTGTGTGTGATCATCAGCACTCAACCTCCCAAGTATCAAGCTGCTTGGAG 540
QY 181 GlnPheAspLeuGlnArgGlyTyrAspThrLeuThrValGlyAspGlyGlnAspGly 200
DB 541 GAGTTGATTTGGAGAGGGGCTATGACACCTGACGCTGCGTATGTGTGAGAGAGGG 600
QY 201 AspGlnLysThrValLeuTyrMetSerGlnAsnAlaCysSerAspSerProHisThrPro 220
DB 601 GACCAAGAGACAGTTCTCTAAGTCTCAAAATGCTGCAAGTACAGACCTTCACACCCA 660

```

Qy 221 GlySerArgIleProGluSerMetSerGlyAspIleThrPakGluIleThrValIleu 240
Db 661 GGCTTCGCATCCAGAGAGCATGTCTGGGACATCTGAGGCGAAGATGACTGTACTT 720
Qy 241 GluIleCyAspArgPheIleSerSerSerAspAlaArgSerGlySerValArgIleSerPro 260
Db 721 GAGATCTGTCTGACATTAGCAGTTGACATGCAAGGTTCAGGTTCAGTGAAGAGTCTCA 780
Qy 261 LysThrSerAspAlaValGluIleuValAlaProGlyThrGluIleGluIleGlySerCys 280
Db 781 AAGCTTCTAATGCTGTGGAACTTGCTCTGGGACAGAGATCGACAGGCGCATGTTGC 840
Qy 281 GlyAspProGlyIleProAlaIleThrGlyArgArgGluIleSerArgPheHisGlyAsp 300
Db 841 GGTAACTTCGACATCTGCAATGCGCGAGGAGAGGCTCCCGGTTTACCAACGGTGCAC 900
Qy 301 ThrLeuIleuPheGluIleGluIleProAlaPheGluIleuValGlyGluIleValIleThrCys 320
Db 901 AACCTCAAGTTGAGTGCACGCCGCTTGAGGTGGTGGGACAGAGCAATCACAATGC 960
Qy 321 GluIleuAspAsnGluIleThrSerAlaIleuValArgIlePheSerCysPhePheAsn 340
Db 961 CAAAGATATACCAATGCTGCGCTAAGAGCAGGCTGCTGCTCTCTCTTCTTCAAC 1020
Qy 341 PheThrSerProSerGlyValValIleuSerProAsnIleThrProGluAspIleGlyAsnHis 360
Db 1021 TTCAACAGCCGCTGCGGATGCTCTGCTCTCTCAATCCAGAGACTATAGGCAACAC 1080
Qy 361 LeuHisCysValIleThrLeuIleuValArgProGluSerArgIleHisIleuValAlaPheAsn 380
Db 1081 CTCACAGTGTCTGGCTCATCTGGCCAGGCTGAGAGCCGCAATCCACCTGGCTTCAAC 1140
Qy 381 AspIleAspValGluProGluPheAspPheIleuValIleuValAspGlyValIleThrAlaGlu 400
Db 1141 GACATTGACGTGGAGCCTCAGTTGATTTCTGTGTCATCAAGATGGGCGCACCGCCAG 1200
Qy 401 AlaProValIleuGlyThrPheSerGlyAsnGluIleuProSerSerIleThrSerSerGly 420
Db 1201 GCGCCGCTGCGGACCTCTCAGAGAAACAGGCTCCCTCTCATACAGCAAGTGGC 1260
Qy 421 HisValAlaArgLeuGluPheGluIleThrAspHisSerThrGlyIleuValArgIlePheAsnIle 440
Db 1261 CACGTGGCCGCTCGAGTTCAGACTGACCACTCCAGAGGAGAGGCGCTTCAACATC 1320
Qy 441 ThrPheThrThrPheArgHisAsnGluCysArgProAspProGlyValProValAsnGlyIle 460
Db 1321 ACTTTTACCACTTCGACACAGACAGAGCGCGGATCTGCGCTTCCAGTAAATGGCAAA 1380
Qy 461 ArgPheGlyAspSerIleuGluIleuGlySerSerIleSerPheIleuCysAspGluIlePhe 480
Db 1381 CGGTTTGGGAGACGCTCAGCTGCGGACGCTCCATCTCTCTGTGATGAAGGCTTC 1440
Qy 481 LeuGlyThrGluIleuGlySerGluThrIleThrCysValIleuIleuGluIleuValIleThr 500
Db 1441 CTTGGAGACTCAGGCTCAGAGACATCACTGCGCTCGGAGAGGAGGAGGAGTGTCTGG 1500
Qy 501 AsnSerAlaValIleuArgCysGluAlaProCysGlyIleHisIleuThrSerProSerGly 520
Db 1501 AACAGCGCTGTGCGGTGTGAAGCTCCCTGTGTGTGTCTCACTGACTCCGCCAGCGGC 1560
Qy 521 ThrIleuSerProGlyIleThrProGlyIleThrIleuAspAlaIleuSerCysAlaIleThrVal 540
Db 1561 ACCATCTCTCTCGGAGCTGCGCTGCTTCAAGAGATGCTTGAGCTGTGCTGGGTG 1620
Qy 541 IleGluAlaGluProGlyIleThrProIleuValIleThrPheAspArgPheIleThrGluIleVal 560
Db 1621 ATTGAGGCCAGCAGGCTACCCATCAAAATCACTTCGACAGATTAACAAACGAGGTC 1680
Qy 561 AsnIleuValIleuGluIleuValArgAspGlyIleThrIleuValIleuIleGlyVal 580
Db 1681 AACATATGACACCTCGAGATGAGCGGATGGCGGACTTACTACAGCGCCCTTATGATGGGGTT 1740
Qy 581 TyrHisGlyThrGluIleuValProGluPheIleuIleSerThrSerAsnIleuIleuIleu 600

Db 1741 TACACGGAGACCAAGTTCCTCCAGTCTCTATAGACACAGCACTACTCTACTCTC 1800
Qy 601 PheSerThrAspIleuSerHisSerAspIleGlyPheGluIleuArgIleGluThrIleThr 620
Db 1801 TTCTCTACCAAGAGATCACTCGACATCGGCTTCCAGCTCCGCTATGAGACTATACA 1860
Qy 621 LeuGluIleuSerAspHisCysIleuAspProGlyIleProValIleuGluIleuArgHisGlyAsn 640
Db 1861 CTGCAATGACACCACTCTGAGATCCAGAGATCCAGATTAATGACAGCGCATGGGAAT 1920
Qy 641 AspPheThrValGlyAlaIleuValIleThrPheSerCysAspSerGlyIleThrIleuSerAsp 660
Db 1921 GACTTCACTGAGGCGGCTGTGTACCTTCACTGTATCTCGGCTACATTAAGTAC 1980
Qy 661 GlyIleuProIleuGluIleuGluIleuProAspPheGluIleuThrArgAlaIleuProSerCysGlu 680
Db 1981 GGGAGGCTCTGGAGTGTGAGCCCACTTCAGAGGAGCGGCGCTGCCAGTTGTGAA 2040
Qy 681 AlaIleuIleuGlyIleuPheIleuGluIleuGlySerSerGlyIleThrIleuSerProGlyIlePhePro 700
Db 2041 GCTCTGTGTGTGCTTCAATTCAGAGCTCCAGTGGACATCTTGTCAGAGGATTCCT 2100
Qy 701 AspPheThrProAspAsnIleuAsnIleuIleuThrIleuIleuGluIleuThrSerHisGlyIleGly 720
Db 2101 GACTTCACTCCCAAGCACTTGAAGTCACTGATTTATCAAAATCTCATGCAAGGT 2160
Qy 721 ValPhePheThrPheHisIleuThrPheHisIleuGluIleuSerGlyIleuValIleuValIleThr 740
Db 2161 GTGTTCTTCACTTCCACACTTCCACTGGAAGTGGCATGACTACTCTCATCATCT 2220
Qy 741 GluAsnGlySerPheThrGluIleuProIleuValGluIleuThrGlySerArgIleuProAlaPro 760
Db 2221 GAGAACGCGAGCTTACCCAGCCCTGAGGACGATCACTGATCTGCGTCCAGCTCC 2280
Qy 761 IleSerAlaGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 780
Db 2281 ATCAGCCCTGGCTCTATGCACTTCACTGCTCCAGGCTCCGCTTCACTGATTTCTCC 2340
Qy 781 MetSerThrGluIleuPheAsnIleuThrPheSerGluIleuIleuIleuIleuIleuIleuIleu 800
Db 2341 ATGTCATATAGAGATTCACATCACTTCCAGAGATGACACTTGGAGCCCTGAGAGAG 2400
Qy 801 ProGluValProAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 820
Db 2401 CCCAGGCTCCAGCTCAGACATCCGAGAGGCTTGGAGTTGGCTGGGCGAGCACTTG 2460
Qy 821 ThrPheSerCysPheProGlyIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 840
Db 2461 ACCTTCTCTGCTTCCCGGAGTACCTGTGAGGAGCACCGCCGATCACTGCTGCGG 2520
Qy 841 GlyArgArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 860
Db 2521 GCGAGAGCGGCGCTGTGAGCTGCTCTGCAAGGAGTGTGCTGAGTGGGAAATCA 2580
Qy 861 ValIleuGlyThrGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 880
Db 2581 GTCAAGAGCACTCAGGCTACTTGTGTCTCCCAACTTCTGTGAATCAATTAACAT 2640
Qy 881 HisGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 900
Db 2641 CATATATGATCTTCACTCAATCCAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 2700
Qy 901 AlaPheGluIleuSerGluIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 920
Db 2701 GCATTGCACTTCCAGAGAGATGCTCTCAAGTTATGATGCAACAACACTCGGCC 2760
Qy 921 ArgIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 940
Db 2761 CGTTGTGTGAGTATTACCATTCGAGATATGAGGAGTGTGATTTGAACAGACATCC 2820
Qy 941 SerSerIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 960

Db	2821	AGCAGCTGTGGCTTGATTTTCATCACTATGCTGAAACACACGAGAGGGCTTTGAAC	2860
Qy	961	HisPheSerSerPheGluLeuIleIysCysGluAspProGlyThrProIysPheGlyTyr	980
Db	2881	CACCTTTCCAGCTTTGATGACTCATCAATATGTGAGAACCCAGAAACCCCAAGTTTGGCTAC	2940
Qy	981	LysValHisAspGluGluHisPheHisAspGlySerSerValSerPheSerCysAspProGly	1000
Db	2941	AAAGTTTCATGATGAAAGTCAATTTTGCAGGAGACCTCGATGCTTTAAGCTGTGACCTCGGA	3000
Qy	1001	TyrSerLeuIysArgGlySerGluGluLeuLeuCysLeuSerGlyGluIleArgThrTyrAsp	1020
Db	3001	TACAGCCTCCGGGGTATAGTAGAGAGCTGTGTCTGATGGAGAGACGCCGGACCTGGGAC	3060
Qy	1021	ArgProLeuProThrCysValHisAspGlyCysGlyValThrValArgGlyGluValSerGly	1040
Db	3061	CGGCTCTGCGCACTGTGTCCCGCAGGTGGAGGAGACAGAGAGAGAGAGTGTCCGGGG	3120
Qy	1041	GluValLeuSerProGlyTyrProAlaProIysGluHisAsnLeuAsnCysIleTyrThr	1060
Db	3121	CAGGTGCTGTCAACCGGGGATTCACACTCCCTATGAACATCTCAATCTGATCTGAGCC	3180
Qy	1061	IleGluAlaGluAlaGlyCysThrIleGlyLeuHisPheLeuValPheAspThrGluGlu	1080
Db	3181	ATCGAAGCAGAGCCGGCCGCTGCACCAATTGGCGTACACTTCCGGTGTGGACACAGAGAG	3240
Qy	1081	ValHisAspValLeuArgIleTyrAspGlyProValGluSerGlyValLeuLeuIysGlu	1100
Db	3241	GTTCAAGACGTCTCTCGCATCTGGGATGGGCTGTGGAGAGCGGGGTTCTCTAAGAGG	3300
Qy	1101	LeuSerGlyProAlaLeuProIysAspLeuHisSerThrPheAsnSerValValLeuGln	1120
Db	3301	CTGAGATGGCCCGCCCTCGCCACAGGACCTGCATAGCACTTCACTCGATCGTCTCGACG	3360
Qy	1121	PheSerThrAspPhePheThrSerTyrGlnGlyPheAlaIleGlnPheSerValSerThr	1140
Db	3361	TTTCACTGACTTCTTTCACAGCAGAGAGGCTTGGCATTCAAATTTTCAGTGTTCACA	3420
Qy	1141	AlaThrSerCysAsnAspProGlyIleProGluAsnGlySerArgSerGlyAspSerTyr	1160
Db	3421	GCAAGCTCTCGCAATGACCTGTGGATCCCGCAGATGGGAGTCCGAGTGGTGAACGTTGG	3480
Qy	1161	GluAlaGlyAspSerThrValPheGlnCysAspProGlyTyrAlaLeuGlnGlySerAla	1180
Db	3481	GAAAGCCGGGACTCCACAGGTGTTCAGGTGACCTGTGGCTACCGGCTGCAGGAGAAAGCA	3540
Qy	1181	GluIleSerCysValIysIleGluAsnArgPhePheTyrGlnProSerProProThrCys	1200
Db	3541	GAGATCACTGTGTGAATCGAAGACAGTTCCTTCTGGCAGCCCAACCCGCAACATGC	3600
Qy	1201	IleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerProAsnTyr	1220
Db	3601	ATGCTCTCCCTCGGGGAGACTTCACAGAACCATCTGGAGTCACTCTCAACCAAAATTAC	3660
Qy	1221	ProGluProTyrProProGlyIysGluCysAspTyrIysValThrValSerProAsnTyr	1240
Db	3661	CCAGAACCTTACCCGCCAGGACGAGAGGTGTACTGGAAAGTCAACCGTCTCAACAGACTAC	3720
Qy	1241	ValIleAlaLeuValPheAsnIlePheAsnLeuGluProGlyTyrAspPheLeuHisIle	1260
Db	3721	GTCATCGCCCTCGATTTAAATCATCTTTAACTCGAGACCTGGCTATGACTTCTCCATATC	3780
Qy	1261	TyrAspGlyIysAspSerLeuSerProLeuIleGlySerPheTyrGlySerGlnLeuPro	1280
Db	3781	TACGACGAGCGGAGTCTCTCAAGCCCTCTCAATAGAAAGCTTCAATGAGCTCCAGCTCCA	3840
Qy	1281	GlyArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspAlaSerVal	1300
Db	3841	GGCCGCATTTGAAAGAGACGACGACAGCAAGCTTCTCGCCCTTCCGACAGCATGATCTGTG	3900
Qy	1301	SerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysPheAspPro	1320
Db	3901	AGCAATGTGGCTTCTGTCATTTAGACTTAAACGAAACCCCGGGAGTCACTATTTTGTATCTCT	3960

QY	1321	GIYSERILEYSAAGCLYTHARGVALGLYSERAPLEULYSLEUGLYSERSEVALYTHR	1340
DB	3961	GGTTCATCAAMAGAGGACACAGGGGTGGGGTCCAGCTCGAAGCTGGGCTCTCCGTACCC	4020
QY	1341	TYTYRCYHISAGLYGLYTYRGIVUALGLIULYTHRSERTHRLSEUSCYEILILEUGLY	1360
DB	4021	TACTACTGCCACGGGGGCGTACGAGTTGAGGGACCTCGACCCTGAGCTGCATCTGGGG	4080
QY	1361	PROAARGLYLVSPPROVALTPRASNAPPROARPPROVALCYETHRALAPROCYGGLYGLY	1380
DB	4081	CCGTATGGGAACCCGGTGGAGCAATCCCGGCGCAGCTGCACAGCCCTCTGGGGGA	4140
QY	1381	GLINTYVALGLYSERASPGLYVALVALLEUSERPROANTYRPROGLINANTYRTHSER	1400
DB	4141	CAGTATGGGGGTTCGAGCGAGTGGTCTTGTCCCCCAACTACCCCGAGACTACACAGT	4200
QY	1401	GLYGLINILECYELEUTYRPHETHALYVALPROLYSAPTYRVALVAPHEGLYGLNPE	1420
DB	4201	GGACGAGACTGCTGTTATTTGTATTACTGTGCCAAGGACTATGTGGTGTGGCCAGTTCC	4260
QY	1421	ALAPHEPHEIETHRALALEUASAPVALVALIULVALHISAPGYHISERGLNHIS	1440
DB	4261	GCTTCCTTTCACAGGCCCCCTCAACGAGCTGTGAGGTTCACGAGCGCCACAGCCACAC	4320
QY	1441	SEARATGLEULSEUSERSEUSERSEGLYSETHISTHRGLYGLSEURLEUPOLEUALATHR	1460
DB	4321	TGCGGGCTCTCAGAGCTCCCTCGGGGCTCCCATACAGAGATACATCGCCCTTGGCCACC	4380
QY	1461	SEARANGINVALLEULILEYSPHSESRALALAGLYLEUALAPROALALARGLYPHNHIS	1480
DB	4381	TCCAAATCAAGTTCATTAAAGTTATGAGGCCAAGAGCCCTCCACACAGCCAAAGGCTTCAC	4440
QY	1481	PHETHALTYRGINALVALAPROARGTHRSERIALATHRGINCYSEUSERSEVALPROGLPRO	1500
DB	4441	TTGTGCTHCCAAAGGGGTTCCCTCGAACACAGCCCAACGAGTGCAGCTCTGTGCCGAAACC	4500
QY	1501	ARGTYRGILYLSARGLEUGLYSERASPHSESERVALGLYALILEVALARCPHEGLUCYS	1520
DB	4501	CGTATGCGCAAGAGCGCTGGCGAGAGCACTTCGCTGGGGGCGCAATCGTCCGCTTCGAATGC	4560
QY	1521	ASNSERGILYTYRALALEUGINGLYSESPROGLINILEGLUCYSEUPROVALPROGLYALA	1540
DB	4561	AACTCCGGCTATGCCCTCGACAGGGGTCCACAGATGCAAGTGCCTCTGTGGCTGGGGCC	4620
QY	1541	LEUALAGINTYRPAENVALSERIALPROTHR CYSEVALVALPROCYSGILYASMLEUTHR	1560
DB	4621	TTGGGCCCATGGAATGCTCAGCGCCACAGTGTGTGTGTGTCGCTGTGGAGCCAACTCACA	4680
QY	1561	GLIUNHPRAGGLYTHRIILEUSERPROGLYPHEPROGLINPROTYRLEUANSERLEUASN	1580
DB	4681	GAGGCGAGGGGCACCAATCTGTCTCCCTGGCTTCCCAAGCGGTACTCAACACACTTCAC	4740
QY	1581	CYSVALTPRPLYSILEVALVALPROGLUGLYALAGLYILEGLINILEGINVALVALSERPHE	1600
DB	4741	TGTGTGGGAAGATCGTGTCTCCGAAAGGGCGTGCATCCAGATCCAAAGTTGCAAGTTT	4800
QY	1601	VALTHRGIVUNASANTPRASPSERLEUGLIVALPHEASPGLYALASAPANTHRVALYTHR	1620
DB	4801	GTCACAGGCGAAGACTGGACTCGCTCGEAGATTTGATGTGTCACATATAACACTGTAAACC	4860
QY	1621	METLEUGLYSERPHESERGLYTHRTTHVALPROALILEUENANSETHRSERANGIN	1640
DB	4861	ATGCTGGGAGTTTCTCAGAAACAAACCGTCCCTGCCCCCTTGTGAACAGCACTCCAAACGAG	4920
QY	1641	LEUTYRLEUHHISPHYRYRSEASPILESERVALSERIALAAGLYPHEHISLEUGLUTYR	1660
DB	4921	CTTACCTTCATTCTTACTCAGATATCAGAGGTATCTGAGCTGCTTCACCTTGGAGTAC	4980
QY	1661	LYETHRVALGLYLEUSERSECYSPROGLINPROALVALPROSERKANGLYVALYVETHR	1680
DB	4981	AAAACGGGTGGCCCTGACCAATGTGTCCGGAACCTGCTGTGCCAGTAAACGGGTGGAAGACT	5040

1691 GlyIuaTgTYrLeuValAsnAspValSerPheGlnCysGluProGlyTYrAlaLeu 1700
5041 GCGAGCGCTACTGTGTAATGATGTGCTTTCAGGTGAGCGGGATATGCCCTC 5100
1701 GlnGlyHISAlaHISISerCysMetProGlyThrValArgArgTrpAsnTYrPro 1720
5101 CAGGGCCACGGCCATCTCCGTGACGCCGGAACAGTGGCGCATGAACTACCTCTC 5160
1721 ProLeuCysIIeAlaGlnCysGlyGlyThrValGluGluMetGluGlyValIIeLeuSer 1740
5161 CCACCTGTATGTGACAGTGTGGGGAAACAGTGAAGAGATGGAGGGGTATCTTGAGC 5220
1741 ProGlyPheProGlyAsnTYrProSerAsnMetAspCysSerTrpLYsIIeAlaLeuPro 1760
5221 CCCGGCTTCCAGGGAATACCCCACTGAATGACTCTCTCTGAAAATAGCACTGCC 5280
1761 ValGlyPheGlyAlaHISISerPheLeuAsnPheSerThrGluProAsnHISAspTYr 1780
5281 GTGGGCTTTGGAGCTCATCATCCAGTTCTTGAACTTCCACCGAGCCCAACAGACTAC 5340
1781 IIeGluIIeArgAsnGlyProTYrGluThrSerArgMetMetGlyArgPheSerGlySer 1800
5341 ATAGAATATCCGAAATGGCCCTTATGAGACAGCCGATGATGGAAATTCAGTGAAGC 5400
1801 GluLeuProSerSerLeuLeuSerThrSerHISGluThrThrValTYrPheHISerAsp 1820
5401 GAGCTTCCAAAGCTCCCTCTCTTCCACGTCCACAGAACACCGGTATTTCCACAGGAC 5460
1821 HISerGlnAsnArgProGlyPheLYsLeuGluTYrGlnAlaTYrGluLeuGlnLYs 1840
5461 CACTCCCAAGATCGGCCAGGATTCAAGCTGAGATCAGGCTTATGAATTCAGAGAGGC 5520
1841 ProAspProGluProPheAlaAsnGlyValIleValArgGlyAlaGlyTYrAsnValGlyGln 1860
5521 CCAAGCCAGAGCCCTTGGCAATGGCAATGTGAGGGAGCTGGCTACAAAGTGGAGAA 5580
1861 SerValThrPheGluCysLeuProGlyTYrGlnLeuThrGlyHISProValLeuThrCys 1880
5581 TCACTGACCTTCAAGTGCCTCCGGGGATATCAATGACTGGCCACCTGTCTTCAAGTGT 5640
1881 GlnHISGlyThrAsnArgAsnTrpAspHISProLeuProLYsCysGluValProCysGly 1900
5641 CAACATGGACCAACCGGAATCGGAGACCAACCCCGCCAAAGTGAAAGTCCCTTGTGC 5700
1901 GlyAsnIIeThrSerSerAsnGlyThrValTYrSerProGlyPheProSerProTYrSer 1920
5701 GGGAAACATCACTTCTTCCAAAGGCACTGTGTAATCCCGGGGTTCCCTAGGCCGTACTCC 5760
1921 SerSerGlnAspCysValTrpLeuIIeThrValProIIeGlyHISGlyValArgLeuAsn 1940
5761 AGCTCCCAAGACTGTGTGTGCTGATATCACCGTCCCATTTGGCCATGGCGTCCGCTCAAC 5820
1941 LeuSerLeuLeuGlnThrGluProSerGlyAspPheIIeThrIIeTrpAspGlyProGln 1960
5821 CTGAGCTGTGTGACAGACAGAGCCCTCTGGAATTTCAATCACCACTGGGAATGGGCAAG 5880
1961 GlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLYsLYsThrValGlnSer 1980
5881 CAACAGACACACGGCTCGGGGTCTTCAACCGAGCATGGCCAAAGAAACGTCAAGT 5940
1981 SerSerAsnGlnValIIeLeuLYsPheHISArgAspAlaIleThrGlyGlyIIePheAla 2000
5941 TCATCCAAACAGTCTGCTCAAGTTCCACCGTATGACAGCCAGGGGGATCTTGCC 6000
2001 IIeAlaPheSerAlaTYrProLeuThrLYsCysProProProThrIIeLeuProAsnAla 2020
6001 ATAGCTTCTCGCTTATCCACTACCAATGACCTCTCTCCACATCTCCCAAGCC 6060
2021 GluValValThrGluAsnGluGluPheAsnIIeGlyAspIIeValArgTYrArgCysLeu 2040
6061 GAATCTGTGACAGAAATGAAGAAATTCATATAGTGAATCGTACATGAGATGCTCT 6120
2041 ProGlyPheThrLeuValGlyAsnGluIIeLeuThrCysLYsLeuGlyThrTYrLeuGln 2060

6121 CTTGGCTTACTAGTGGGGAATGAATTTCTGACCTGCAACTGGAACTTACTGAG 6180
2061 PheGluGlyProProProIIeCysGluValHISCysProThrAsnGluLeuLeuThrAsp 2080
6181 TTTGAAGACCAACCCCGATATGTAAGTCACTGTCAACAAATGAGCTTCTGACAGAC 6240
2081 SerThrGlyValIIeLeuSerGlnSerTYrProGlySerTYrProGlnPheGlnThrCys 2100
6241 TCCAGAGCGTGAATCTGAGCCAGAGCTACCTGGAAGCTATCCCAAGTTCAGACTGC 6300
2101 SerTrpLeuValArgValGluProAspTYrAsnIIeSerLeuThrValGluTYrPheLeu 2120
6301 TCTTGGCTGTGAGAGTGAAGCCCGACTATATACATCTCCCTCAAGTGAATCTTCTC 6360
2121 SerGluLYsGlnTYrAspGluPheGluIIePheAspGlyProSerGlyGlnSerProLeu 2140
6361 AGCAGAGAAGCAATATGATGATTGAAATTTTATGATGCTCATCAGAGACAGATCCTCG 6420
2141 LeuLYsValIleLeuSerGlyAsnTYrSerAlaProLeuIIeValThrSerSerAsnSer 2160
6421 CTGAAGCCCTCAAGTGGAAATTAATCACTCCCTGATGTGCACACACTCAAGCAACTCT 6480
2161 ValTYrLeuArgTrpSerSerAspHISAlaTYrAsnArgLYsGlyPheLYsIIeArgTYr 2180
6481 GTGTACTGCTGTGCTCATCTGATCAAGCCCTACAAATGGAAAGGCTTCAAGATCCGCTAT 6540
2181 SerAlaProTYrCysSerLeuProArgAlaProLeuHISGlyPheIIeLeuGlyGlnThr 2200
6541 TCAGCCCTTATGCAACCTGCCCCAGGGCTTCACTCATGAGCTTCACTTCAAGCAAGC 6600
2201 SerThrGlnProGlyLYsSerIIeHISPheGlyCysAsnHISGlyTYrArgLeuValGly 2220
6601 AGCACAGCCCGGGGCTCCATCACTGAGTGGCTGCAACCGCGCTACCCCTGTGTGGA 6660
2221 HISerMetAlaIleCysTYrThrArgHISProGlnGlyTYrHISLeuTrpSerGluAlaIle 2240
6661 CACAGCATGGCCATCTTACCCGGCACCCCAAGGCTTACCACTGTGAGAGCAAGCATC 6720
2241 ProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLYsAsnGlyMetValPhe 2260
6721 CCTCTGTGAAGCTCTTCTCTGTGGGCTTCTGAGGCCCCCAAGATGGAATGTGTGT 6780
2261 GlyLYsGluTYrThrValGlyThrLYsAlaValTYrSerCysSerGlnGlyTYrHISLeu 2280
6781 GGCAGGAGTACACAGTGGAAACCAAGCCATGTACAGCTGCAAGTGAAGCTTACACTC 6840
2281 GlnAlaGlyAlaGluAlaThrAlaGluCysLeuAspThrGlyLeuTrpSerAsnArgAsn 2300
6841 CAGGACAGCCGTAGGCACTGCAAGATGTCTGAGACACAGGCTTATGAGCAACCGCAT 6900
2301 ValProProGlnCysVal-----ProValThrCysPro 2311
6901 GTCCCAACACAGTGTGCTCGTGAAGTCTCGGGCAATGAGAGCGGGTGTGACTTGTCT 6960
2312 AspValSerSerIIeSerValGluHISGlyArgTrpArgLeuIIePheGluThrGlnTYr 2331
6961 GATGTCAGTGACATCAAGCTGAGACATGGCCGATGAGGGCTTATCTTTTGAGACACAGTAT 7020
2332 GlnPheGlnAlaGlnLeuMetLeuIleCysAspProGlyTYrTYrTYrThrGlyGlnArg 2351
7021 CAGTTCAGGCCCAAGCTGATGCTCATCTGTGACCTTGCTATCTATCTATCTGCGCAAGG 7080
2352 ValIIeArgCysGlnAlaAsnGlyLYsTrpSerLeuGlyAspSerThrProThrCysArg 2371
7081 GTATCCGCTGTACAGGCAATGGCAATGAGAGCTCGGGAGCTTACGCCCACTGCGCA 7140
2372 IIeIleSerCysGlyGluLeuProIIeProProAsnGlyHISArgIIeGlyThrLeuSer 2391
7141 ATATCTCTCGTGAAGGCTCCGATTTCCCCCAATGGCCACCGCATGGAAACATGTCT 7200
2392 ValTYrGlyAlaThrAlaIlePheSerCysAsnSerGlyTYrThrLeuValGlySerArg 2411

1018 hrtPaAspArgProleuProThrCysValAlaGluCysGlyGlyThrValArgGlyGluV 1038
549 CCGGGAGACGGGCTCTGCCCACTGTGTCCGCACTGGAGGAGACATGAGAGAGAGG 608
1038 aLSerGlyGluValLeuSerProGlyTyrProAlaProTyrGluHisAsnLeuAsnCysI 1058
609 TGTGGGGGACAGTGTGTGACCCGGGGATCCAGCTCCCTATGAAACAATCTCAACTGCA 668
1058 leTPrThrIleGluValaGluValaGlyCysThrIleGlyLeuHisPheLeuValPheAspT 1078
669 TCTGACCATGGAAGACAGAGCCGGCTGCACATTTGGGCTACACTTCTGTGTGTTTACA 728
1078 hrGluGluValHisAspValLeuArgIleTPaAspGlyProValGluSerGlyValLeuL 1098
729 CAGAGAGAGTTCAAGACGTGTGGGATCTGGGATGGGCTGTGGAGAGCGGGGTTCTGC 788
1098 euLyGluLeuSerGlyProAlaLeuProLyAspLeuHisSerThrPheAsnSerValV 1118
789 TGAAGAGAGTGAAGTGGCCCGGCTGGCCCAAGAGACTGTACATAGCACTTCAACTCGGTGC 848
1118 alleGluInPheSerThrAspPheThrSerIleGlyGlyPheAlaIleGluInPheSerV 1138
849 TCTGCAAGTTCAAGACATGACTTCTTTCACACAGCAAGCGGGCTTGGCCATTCATTTTCAG 908
1138 aLSerThrAlaThrSerCysAsnAspProGlyIleProGluHisAsnGlySerArgSerGlyA 1158
909 TGTCCACAGCAAGACTCTCTGCAATATACCTGGGATTCGCGAGAAAGGGAGTGGAGTGTG 968
1158 apSerTPGluValaGlyAspSerThrValPheGluCysAspProGlyTyrAlaLeuGluG 1178
969 ACAGTGGGAGAGCGGGGAGCTCCACAGTGTCCAGTGTGACCTGGGCTGAGCGGTGCAGG 1028
1178 lySerAlaGluIleSerCysValIySileGluAsnArgPhePheTyrGluInProSerProp 1198
1029 GAAATGCAAGATCAGCTGTGTGAAGATCGAGAAACAGGTCTTCTGCGAGCGCCCGCGC 1088
1198 roTPaCysIleAlaProCysGlyGlyAspLeuThrGlyProSerGlyValIleLeuSerP 1218
1089 CAACATGATGTGCTCTCCGCGGGGAGACCTGACAGACATCTGGAGTCACTCTCTCAC 1148
1218 roAsnTyrProGluInProTyrProProGlyIySileGluCysAspTyrLyValThrValSerP 1238
1149 CAATTTACCCAGAACCTTACCCTGCGAGCAAGAGTGTGAGTGAAGAGTACCGTCTCAC 1208
1238 roAspTyrValIleAlaLeuValPheAsnIlePheAsnLeuGluInProGlyTyrAspPheL 1258
1209 CAGACTACGTATCGCCCTGTATTATTAACATCTTAACTGAGACCTGGCTATGACTTCC 1268
1258 euHisIleTyrAspGlyAArgAspSerLeuSerProLeuIleGlySerPheTyrGlySerG 1278
1269 TCCATATCTTACAGAGAGAGGAGCTCTCTCAGCCCTCTCATAGAGGCTTCTATGGCTCC 1328
1278 InLeuProGlyAArgIleGluSerSerSerAsnSerLeuPheLeuAlaPheArgSerAspA 1298
1329 AGCTCCCAAGGCGGATTTGAAGAGCAAGCAACAGCTTCTCTGCTTCCGAGCGATG 1388
1298 laSerValSerAsnAlaGlyPheValIleAspTyrThrGluAsnProArgGluSerCysP 1318
1389 CATCTGTAGCAATGCTGGCTTCTGTATGACTTACAGAAACCCCGGAGAGCATGTT 1448
1318 heAspProGlySerIleLyAsnGlyThrArgValaGlySerAspLeuLyLeuGlySerS 1338
1449 TTGATCTCTGTTTCATCAAGAGCGGACACCGGGTGGGCTCGACTTAAGCTGGGCTCT 1508
1338 erValIThrTyrTyrCysHisGlyGlyTyrGluValaGluGlyThrSerThrLeuSerCysI 1358
1509 CCGTCACTACTACTGCGACAGGGGGCTTACGAAGTTAGGGGCACTCGACCTGAGCTGCA 1568
1358 leLeuGlyProAspGlyLyAspProValITrpaAsnAspProArgProValCysThrAlaProC 1378
1569 TCTGGGGGCTGTAGTGGAGAGCCGCTGTGGAACAATCCCGGCGAGTCTGCAAGACCCCT 1628

1378 ySGLyGlyInTyrValaGlySerAspGlyValaValLeuSerProAsnTyrProGluAsnT 1398
1629 GTGGGGACAGATATGTGGTTCGAGCGAGATGATCTTGTCTCCCAATACCCAGAACT 1688
1398 yThrSerGlyGluIleCysLeuTyrPheValIThrValIPolyAspTyrValaIlePheG 1418
1689 ACACAGTGAACAGATGTCTGTATTTTGTACTGAGCCCAAGACATATGTGTTGTTG 1748
1418 lyGluPheAlaPhePheHisThrAlaLeuAsnAspValaValaGluValHisAspGlyHis 1438
1749 GCCAGTTCGCTCTTTCATACAGCGCCCTCAACAGCGTGTGGAGTTTCAGAGCGGCACA 1808
1438 erGluHisSerArgLeuLeuSerSerLeuSerGlySerHisThr----- 1452
1809 GCCAGCATCGCGGCTCTAGCTCCTCTCGGGCTCCCTAACAGATATCCGGGGCTCGG 1868
1452 ----- 1452
1869 CCAAGTGGGGATGTGTGGCCGGGGGATCATCGTCCGGTTAAAGAAAGAGGCTCTA 1928
1452 ----- 1452
1929 GAAGCAACCCATGCGCGCAGGTGAAACCTACGGCTTGTGTCGTGTGTTGTG 1988
1452 ----- 1452
1989 CTTGTCTCAACCGCAGACGACAGCTGTAGAGCTCCAACTAGCGGGGCTTACAGACGT 2048
1452 ----- 1452
2049 GCCCTCACCGAGATGTGTACACCGCCCTTGTGTGTAGCTTCTCTGTGTAATG 2108
1452 ----- 1452
2109 GCAACTACATAATGTGCTGACAGTCCAGTGTGTCTCTCCCTGGCCCATGTGTA 2168
1452 ----- 1452
2169 CTGCAACCAAGACAGATATACCTTGTCTTCTGTCTCAAAAGCTGTCACTACCTCG 2228
1453 ----- 1453
2229 TTTCTGTGCCCATGACAGAGATCACTGCTTGTGCACTCAATCAAGTTCTCTATTA 2288
1467 yPheSerAlaLySileGlyLeuAlaProAlaArgGlyPheHisPheValITyrGlu----- 1484
2289 AGTTCAAGCGCAAGAGGCTTCGACACGCGCAGAGCTTCCACTTGTCTACCAAGTATG 2348
1485 -----AlaValProArgThrSerAlaThrGluCysSerSerValProG 1499
2349 AGACATGACGCGCGAGCGGATCTCTGAACAGCGCCACAGCGAGTGCAGCTGTGCGG 2408
1499 InProArgTyrGlyLyAspArgLeuGlySerAspPheSerValaValaIleValaIArgPheG 1519
2409 AACCCCGTATGCAAGAGCTGGGAGTATCTTCTCGGTGGGCGCATGTCCGCTCG 2468
1519 lUcYAsnSerGlyTyrAlaLeuGluGlySerProGluIleGluCysLeuProValProG 1539
2469 AATGCAACTCCGGCTATGCTCTCAGAGGGTCCGACAGATCGAGTGTCTCTCGTGTGCTG 2528
1539 lyAlaLeuAlaGluInTPaAsnValSerAlaProThrCysValaValaProCysGlyAsnL 1559
2529 GGAGCTTGGCCCATAGGAATGTCTCAGCGCCACAGTGTGGTGGCGGTGGAGGCAAC 2588
1559 euThrGluArgArgGlyThrIleLeuSerProGlyPheProGluProTyrLeuAsnSerL 1579
2589 TCAAGAGGCGCAGGGGACATCTGTCCCTGCTTCCAGAGCGGTACTCAACAGCC 2648
1579 euAsnCysValITrplySileValaValaProGluGlyValaGlyIleGluIleGluValaValS 1599
2649 TCAACTGTGTGTGAAGATGTGTGTCCCGAAGCGCTGCAATCCAGATTCAGATGTGTA 2708
1599 erPheValIThrGluGluAsnTrpAspSerLeuGluValaPheAspGlyAlaAspAsnThrV 1619

D	2709	GTITTTGTCAGACAGACAGAACTGGGACTCGCTGAGAAATTGAGTGACGATTAACACTG	2768	D	3789	GCCCGTACTCCAGCTCCAGAGACTGTGCTGATCAACCGTGCCCAATTGGCCATGG	3848
Q	1619	alThrMetLeuGlySerPheSerGlyThrValProAlaLeuLeuAsnSerThrSerA	1639	Q	1936	lyValArgLeuLeuLeuSerLeuLeuGlnThrGluProSerGlyAspPheIleThrIleT	1956
D	2769	TAAACAATGCTGGGAGTTTCTCAGAGAACACCGTGGCTTCTTCAACAGCACTCCCA	2828	D	3849	GGCTCGGCTCAACCTCAGCTGTGTCAGACAGAGCCCTCTGAGATTTTCATCACCATT	3908
Q	1639	snGlnLeuTYrLeuHisPheTYrSerAspIleSerValSerAlaAlaGlyPheHisLeuG	1659	Q	1956	rPAspGlyProGlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaLeuL	1976
D	2829	ACCAAGCTTACTTCTCATTTTCTACTCAGATNACAGATNACGATGCGAGCTGGCTTCCACTGG	2888	D	3909	GGGATGGGCAACAGCAACAGCAACAGCCGCTGGGCTCTTCAACCGGAGCAATGGCCAA	3968
Q	1659	luTYrIleThrValGlyLeuSerSerCysProGluProAlaValProSerAsnGlyValL	1679	D	1976	ysrThValGlnSerSerSerSerAsnGlnValLeuLeuLysPheHisArgAspAlaAlaThrG	1996
D	2889	AGTCAAAACGCTGGGCTGAGCACTTCCGGAACCTGCTGTGCCAGTAAACCGGGGTGA	2948	D	3969	AAACAGTGAAGATTTCATCAACAGAGTCTGCTCAAGTTCCACCGTGATGACGACAG	4028
Q	1679	ysrThrGlyValArgTYrLeuValAsnAspValValSerPheGlnCysGluProGlyTYrA	1699	Q	1996	lyGlyIlePheAlaIleAlaPheSerAlaTYrProLeuThrLysCysProProThrI	2016
D	2949	AGACTGGGAGCGCTTACTTGTGTGATGATGTGTGTCTTCCAGTGTGACCCGGGATATG	3008	D	4029	GGGAGATCTTCCCAAGCTTCTCGCTTATTCCTCACTCAACAAATGCCCTCTCCACCA	4088
Q	1699	laLeuGlnGlyHisAlaHisIleSerCysMetProGlyTYrThrValArgArgTYrAsnTYrP	1719	D	2016	leLeuProAsnAlaGluValValThrGluAsnGluGluPheAsnIleGlyAspIleValA	2036
D	3009	CCCTCCAGGGGCAAGCCCAATCTCTGCANTGCCGGAACAGTGGCCGATGGAACTGACC	3068	D	4089	TCCTCCCAAGCCGGAAGTCGTCAAGAGAAATGAAGATTCAATTAGTGACATCTTAC	4148
Q	1719	roP-roP-roLeuCysIleAlaGlnCysGlyGlyTYrThrValGluGluMetGluGlyValIleL	1739	Q	2036	rgTYrArgCysLeuProGlyPheThrLeuValGlyAsnGluIleLeuThrCysLysLeuG	2056
D	3069	CTCTCCACTCTGTATTCACAGTGTGGGGAAACAGTGAAGATGAGAGGGGTGATCC	3128	D	4149	GCTACAGATGCTCTCGCTTACCTTAAGTGGGAATGAATTCCTGACCTGCAAACTTG	4208
Q	1739	euserProGlyPheProGlyAsnTYrProSerAsnMetAspCysSerTYrIleAlaAla	1759	Q	2056	lyTYrThrLeuGlnPheGlnGlyProProProIleCysGluValHisCysProThrArg	2076
D	3129	TGAGCCCCGGCTTCCAGGCACTACCCCAAGTAAATGACTGTCTCTGAAATAATGACAC	3188	D	4209	GAACCTACCTGAGTTGAGAGCAACCCGATATGAGTGAAGTGCATGTCACAAATAG	4268
Q	1759	euproValGlyPheGlyValHisIleGlnPheLeuAsnPheSerThrGluProAsnHisA	1779	D	2076	luLeuLeuThrAspSerThrGlyValIleLeuSerGlnSerTYrProGlySerTYrProG	2096
D	3189	TGCCCGTGGGCTTGGAGCTCACATCCAGTTCCTGAACTTCTCCACCGAGCCCAACACAG	3248	D	4269	AGCTTCTGACAGACTCCACAGGCGTATCTGAGCCAGGCAAGCTACCTGGAAGCTATCCC	4328
Q	1779	spTYrIleGluIleArgAsnGlyProTYrGluThrSerArgMetGlyArgPheSerG	1799	Q	2096	lnPheGlnThrCysSerThrLeuValArgValGluProAspTYrAsnIleSerLeuThrV	2116
D	3249	ACTACATAGAATAATCCGAATGGCCCTTATAGACACACCCGATATGAGGAAGTTACAGTG	3308	D	4329	AGTTCCAGACCTGCTTGGCTGTGGAGAGTGAAGCCGAGCATATACATCTCCCTCACAG	4388
Q	1799	lySerGlnLeuProSerSerLeuLeuSerThrSerHisGluThrThrValTYrPheHisS	1819	Q	2116	alGluTYrPheLeuSerGlnLysGlnTYrAspGluPheGluIlePheAspGlyProSerG	2136
D	3309	GAAGCGAGCTTCCAAAGTCTCTCTCTCCAGTCCACGAGACACACCGTATTTTCCACA	3368	D	4389	TGGAGTACTTCTTCCAGCGAAGAACATATGATGATGAGTTTGAATTTTATGCTCATCG	4448
Q	1819	erAspHisSerGlnAsnArgProGlyPheLysLeuGlnTYrGln-----	1833	D	2136	lyGlnSerProLeuLeuLysAlaLeuSerGlyAsnTYrSerAlaProLeuIleValThrs	2156
D	3369	GCGACCACTCCCAAGATCGGCCAGAGATTCAAGCTGAGATTCAGATTGACTTACTCC	3428	D	4449	GACAGAGTCTCTGCTGTAAGGCCCTCAGTGGGAATTATCTCAGCTCCCTGATTTGCACA	4508
Q	1833	-----	1833	Q	2156	erSerSerAsnSerValTYrLeuArgTYrPserSerAspHisAlaTYrAsnArgLysGlyP	2176
D	3429	ACCAAGATTCTCTCTCTGAGAGGTTTGATCTCTCGAGTTGAAAGAACCAACTCA	3488	D	4509	GCTCAAGCAACTGTGTATCACTGCGTGTGATCATGATCAGCGCTTACAAATCGGAAGGCT	4568
Q	1834	-----	1837	Q	2176	heLysIleArgTYrSerAlaProTYrCysSerLeuProArgAlaProLeuHisGlyPheI	2196
D	3489	CTCTCCCGCTGCGCGCTTCTATGTCGAGATCTTGATCTTGATCTGTGTAAGCTTATGAA	3548	D	4569	TCAAGATCCGCTATTCA-----	4585
Q	1837	euglnGluCysProAspProGluProPheAlaAsnGlyIleValArgGlyAlaGlyTYrA	1857	Q	2196	leLeuGlyGlnThrSerThrGlnProGlyLysIleHisPheGlyCysAsnAlaGlyT	2216
D	3549	TTCAAGAGTGCACAGACCCGAGAGCCCTTTCCAATGGCATTTGAGAGGGAGCTGCTACA	3608	D	4586	-----GCGCAGACGACGACCCAGCCGGGGCTCATCATCATTTGGCTGCAACCGCGCT	4640
Q	1857	snValGlyGlnSerValThrPheGluCysLeuProGlyTYrGlnLeuThrGlyHisPheV	1877	Q	2216	yrArgLeuValGlyHisSerMetAlaIleCysThrArgHisPProGlnGlyTYrHisLeuT	2236
D	3609	ACGTGGGACATATAGTACTTGAGTGCCTCCCGGGGTATCATATGACTGGGCAACCCG	3668	D	4641	ACCGCTGTGGGACACAGATGCGCATCTGTATCCCGGCAACCCGAGGCTTACCACTGT	4700
Q	1877	alLeuThrCysGlnHisGlyTYrAsnArgAsnTYrAspHisPProLeuProLysCysGluV	1897	Q	2236	rPserGluAlaIleProLeuCysGlnAlaLeuSerCysGlyLeuProGluAlaProLysA	2256
D	3669	TCCTCAAGTGTCAACATGGACCAACCGGAACGGAACACCCCTGCGCCCAAGTGTGAAG	3728	D	4701	GGAGGAAGCATTCCTTCTGTCAAGCTTCTTCTGTGGGCTTCTCTGAGGCCCCCAAGA	4760
Q	1897	alProCysGlyGlyAsnIleThrSerSerAsnGlyTYrValTYrSerProGlyPheProS	1917	Q	2256	snGlyMetValPheGlyLysGluTYrThrValGlyTYrLysAlaValTYrSerCysSerG	2276
D	3729	TCCCTGTGGCGGGAACATCATCTTCCAAACGCACTGTGTCTCCCGGGGGTTCCCTA	3788	D	4761	ATGGAATGTGTGGCAAGAGATACACAGTGGGAACCAAGGCATGTACGTGCACTG	4820
Q	1917	erProTYrSerSerSerGlnAspCysValTYrLeuIleThrVal--ProIleGlyHisG	1936	Q	2276	luGlyTYrHisLeuGlnAlaGlyValAlaThrAlaGluCysLeuAspThrGlyLeuT	2296
D		-----		D	4821	AAGCTACCACTCCAGGCAAGCGCTGAGGCCACTGACGATGTCTGACACAGGCTTAT	4880

QY 2296 rPserAenrAgnValProProglInCyVal-----P 2307
 Db 4881 GGAGCAACCGCAATGTCCCAACAGTGTCTCCGAGTCTCCGGGCAATGAGCGGGT 4940
 QY 2307 rovalThrCyAProApsValSerSerIleSerValGluHisGlyArgTrpArgLeuIleP 2327
 Db 4941 CTGTACCTGTCTCGATGTAGTACATCAGCTGAGACATGCGCGAATGAGGCTTATCT 5000
 QY 2327 hegluThrGlnTyrGlnPheGlnAlaGlnLeuMetLeuIleCyAAspProGlyTyrTyrT 2347
 Db 5001 TTGAGACACAGTATCAGTTCAGGCGCCAGCTGATGCTCATGTGACCTTGCCATCACT 5060
 QY 2347 YrTrnGlyGlnArgValIleArgCySglnAlaGlnGlyLeuTrpSerLeuGlyAspSerT 2367
 Db 5061 ATACTGGCCAAAGGCTATCCGCTGTCCAGGCCAATGAGACCTCGGGAGCTCTA 5120
 QY 2367 hrProthrCyArgIleIleSerCySglnLeuProIleProProAenGlyHisArgI 2387
 Db 5121 CGCCCACTGCGAATCATCTCTGTGAGAGCTCCGATTCGCCCAATGGCCACCGCA 5180
 QY 2387 leglyThrLeuSerValTyrGlyAlaThrAlaIlePheSerCyAAsnSerGlyTyrThrL 2407
 Db 5181 TCGAACAAGTGTCTGTCTACGGGGCAGCAGCATCTTCTCCGCAATCCGATATACAC 5240
 QY 2407 euValGlySerArgValArgIuCyMetAlaAenGlyLeuTrpSerGlySerGlyValA 2427
 Db 5241 TGTGTGGCTCCAGGGTGGAGTGCATGAGCCCAATGGGCTGTGAGTGGCTTGAGATCC 5300
 QY 2427 rGyCyLeu----- 2429
 Db 5301 GCTGCTTGGCACTCAGACCAAGCTCCATCTTCTATAGCTCTCTGATGATAC 5360
 QY 2430 -----AlaGlyHisCySglnThrProGluProIleValAenG 2442
 Db 5361 TCTCTTCCCATCTCCCTCAACCAAGCTGAGACCTGTGGAGCTCCGAGCCCATTTGTCA 5420
 QY 2442 LysHisIleAenGlyGluAenTyrSerTyrArgGlySerValValTyrGlnCyAAsnIleA 2462
 Db 5421 GACACATCAATGGGAGAACTACAGCTACCGGGGAGTGTGTGATCCAAATGCAATGTGTG 5480
 QY 2462 LysPheArgLeuIleGlyMetSerValArgIleCySglnGlnAspHisHisIleTrpSerGlyL 2482
 Db 5481 GCTTCCGCTGATGGGCAATGTCTGTGGCATCTGCGACAGATCATACGTGTGGGCA 5540
 QY 2482 YrThrProPheCyVal----- 2487
 Db 5541 AGACCCCTTCTGTGTGATGTAAAGCAGATTGCTGCTGCTGCTGTGTGTG 5600
 QY 2488 -----ProIleThrCyGlyHisProGlyAenP 2497
 Db 5601 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5660
 QY 2497 rovalAenGlyLeuThrGlnGlyAenGlnPheAenLeuAAspValValLysPheValC 2517
 Db 5661 CTGTCAAGGCGCTCACTAGGGTAAACAGTTTAACCTAAAGATGTGTCAAGTTTGT 5720
 QY 2517 YsaenProGlyTyrMetAlaGlnGlyAlaAlaArgSerGlnCySglnAlaSerGlyGlnT 2537
 Db 5721 GCAACCTGGGATATGTGCTAGGGGCTGCTAGGTCCCAATGCTGCGCAGCGGCAAT 5780
 QY 2537 rPserApsMetLeuProThrCyAArgIleIleAenCyThrAAspProGlyHisGlnGlnL 2557
 Db 5781 GAGAGTGAATGTGCTCCAGCTGCAAAATCATCAATGTACAGATCTTGGAACCAAGAA 5840
 QY 2557 anSerValArgGlnValHisAlaSerGlyProHisArgPheSerPheGlyThrThrValS 2577
 Db 5841 ATAGTGTCTCAGGTCCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5900
 QY 2577 erTyrArgCyAAsnHisGlyPheTyrLeuLeuGlyThrProValLeuSerCySglnGlnL 2597
 Db 5901 CTACCGGTGCAACCAAGGCTTCTACTCTGTGGGCAACCCAGTGTCTGCTGCGGAG 5960

QY 2597 spGlyThrTrpAspArgProArgProGlnCySleu 2608
 Db 5961 ATGGCAATGGAGACCGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5995
 RESULT 4
 US-10-276-934-5
 : Sequence 5, Application US/10276934
 : Publication No. US20030180750A1
 : GENERAL INFORMATION:
 : APPLICANT: University of Leeds
 : APPLICANT: Markham, Alexander F.
 : APPLICANT: Jackson, Andrew P.
 : APPLICANT: Woods, Christopher G.
 : TITLE OF INVENTION: Treatment of Cancer and Neurological Diseases
 : FILE REFERENCE: 9052-144
 : CURRENT APPLICATION NUMBER: US/10/276,934
 : PRIOR FILING DATE: 2002-11-20
 : PRIOR APPLICATION NUMBER: PCT/GB01/02240
 : PRIOR FILING DATE: 2001-05-21
 : PRIOR APPLICATION NUMBER: GB0012186.3
 : PRIOR FILING DATE: 2000-05-20
 : NUMBER OF SEQ ID NOS: 16
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 5
 : LENGTH: 7323
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (34)..(34)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (128)..(128)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (638)..(638)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1502)..(1502)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1538)..(1538)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1586)..(1586)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1782)..(1782)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (1787)..(1787)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (2195)..(2195)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (2213)..(2213)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature
 : LOCATION: (3179)..(3179)
 : OTHER INFORMATION: "n" is any nucleotide
 : FEATURE:
 : NAME/KEY: misc_feature


```

Db 4959 GGGGCTGATGGGAACCCCTCTGGGACCAAGTCGCTCTTCGATGCTCCCTGGGA 5018
Qy 1380 GlyIIntYrValGlySerAspGlyValValLeuSerProAsnYrProGlnAsnYrThr 1399
Db 5019 GGGCAGTACACGGGATCAGAAAGGGTATGTTTATCACCAAACTAACCCCAATATTCACA 5078
Qy 1400 SerGlyIntLeuSerLeuYrPheValThrValProIysAspYrValValPheGlyGln 1419
Db 5079 GCTGTCATAATATGCTCTATTCATTCACGCTACCAAGGAATTCGTCGCTTTGGACAG 5138
Qy 1420 PheAlaPhePheHsThrAlaLeuAspValValGluValHsAspGlyHsSerGln 1439
Db 5139 TTTGCCATATTTCCAGACAGCCCTGAATGATTTGGCAAAATTAATTGATGGAACCATGCA 5198
Qy 1440 HisSerArgLeuSerSerLeuSerGlySerHisThrGlyGlnSerLeuProAlaVal 1459
Db 5199 CAGCCGACCTTCTCACTCACTCTCGGGCTCCTACAGGGGAAACATTCGCCCTTGCT 5258
Qy 1460 ThrSerAsnGlnValLeuIleYsPheSerAlaYsGlyLeuAlaProAlaArgGlyPhe 1479
Db 5259 ACGTCAATCAAAATTCGTCTCCGATTCAGTCAAAAGCGGTGCTCTGCGCGGCTTC 5318
Qy 1480 HisPheValYrGlnAlaValProArgThrSerAlaThrGlnYsSerSerValProGln 1499
Db 5319 CACTTCGTATACAGCTGTCTCTGTACACAGTACACCCAAATGCACTCTGTCCCGAG 5378
Qy 1500 ProArgYrGlyYsArgLeuGlySerAspPheSerValGlyValAlaIleValArgPheGln 1519
Db 5379 CCAGATACGGAAGGAATTTGGTCTTGAGTTCCTGCGCGCTCCACATCGTCCATTCGAG 5438
Qy 1520 CysAsnSerGlyYrAlaIleuGlnGlySerProGlnIleGlnCysLeuProValProGly 1539
Db 5439 TGCACCCGGGATACCTGCTTCAAGGGTTCACAGCGCGCTCCACGTCGCGCGCCAC 5498
Qy 1540 AlaLeuAlaGlnThrPheAsnValSerAlaProThrCysValValProCysGlyGlyAsnLeu 1559
Db 5499 GCTTGGCAGAGTGAACGACAGATCCCGACGTGTGTGTACCTCGCAGTGGCAATTTTC 5558
Qy 1560 ThrGlnArgArgGlyThrIleLeuSerProGlyPheProGlnProYrLeuAsnSerLeu 1579
Db 5559 ACTCAAGAAAGAGTACATCTCTGCTCCCGGCTACCTGAGCCATACGGAACAACCTTG 5618
Qy 1580 AsnCysValTrpYsIleValValProGlnGlyAlaGlyIleGlnIleGlnValValSer 1599
Db 5619 AACTGATATGAAAGATCATATGTTACGAGGGCTCGGGAATTCAGATTCAGAAAGATCACT 5678
Qy 1600 PheValThrGlnIleuAsnThrPheSerLeuGlnValPheAspGlyAlaAspAsnThrVal 1619
Db 5679 TTTGCCACGAGGACGAACTGGAGCTCCCTTGAGATCCAGATGTGGGATGTGGACCGCA 5738
Qy 1620 ThrMetLeuGlySerPheSerGlyThrThrValProAlaLeuLeuAsnSerThrSerAsn 1639
Db 5739 CCAGAGCTGGAGAGCTTCTCAGSCACACAGTACCGGCACTGCTGAACAGTACTTCAAC 5798
Qy 1640 GlnLeuYrLeuHisPheYrSerAspIleSerValSerAlaAlaGlyPheHisLeuGln 1659
Db 5799 CAACTTACCTGATTTCCAGTCTGACATTAAGTGGACAGCTGCTGTTTCCACCTGGA 5858
Qy 1660 TyrLeuThrValGlyLeuSerSerCysProGlnProAlaValProSerAsnGlyValYs 1679
Db 5859 TACAAAACGTGAGGTCTTGTGTCATGCCAAGAACCGCCCTCCCGACGAACGACATCAA 5918
Qy 1680 ThrGlyValArgYrLeuValAsnAspValValSerPheGlnCysGlyProGlyYrAla 1699
Db 5919 ATCGAGATTCGATKATGTGGAAACGACGTCTCTCTTCAAGTGGACGCGGGTACAC 5978
Qy 1700 LeuGlnGlyHisAlaHisIleSerCysMetProGlyThrValArgArgTrpAsnYrPro 1719
Db 5979 CTGCAAGGCGGTTCACCATTTCTGTATAGCAGGAGACCGTTCGCGGTGGAAATATCG 6038
Qy 1720 ProProLeuCysTrpLeuAlaGlnCysGlyGlyThrValGluGluMetGlnGlyValIleLeu 1739
Db 6039 TCTCCCTGTGATTCACCTGTGGAGGACGCTGAGCACTTGGTGTGTGTATCTCG 6098

```

```

Qy 1740 SerProGlyPheProGlyAsnYrProSerAsnMetAspCysSerTrpYsIleAlaLeu 1759
Db 6099 AGCCCCGGCTTCCAGGTTCCTTACCCCAACCTTAAGCTGACCTGGAGATCTCATTA 6158
Qy 1760 ProValGlyPheGlyValHisIleGlnPheLeuAsnPheSerThrGlnProAsnHisAsp 1779
Db 6159 CCCATGGGTATAGTGCATATTCAGTTTCGAATTTTTCACGAAGCTAATCATGAC 6218
Qy 1780 TyrIleGlnIleArgAsnGlyProYrGlnThrSerArgMetCysGlyArgPheSerGly 1799
Db 6219 TTCCTTGAATCAAAATGACCTTACCAACACAGCCCATGATGGAATTTAGCGGC 6278
Qy 1800 SerGlnLeuProSerSerLeuSerSerHisThrIleGlnThrValYrPheHisSer 1819
Db 6279 ACGGATCTCCCGGGCCCTGTGTGACACACAGCATGAAACCTTCATCCTTATATAGT 6338
Qy 1820 AspHisSerGlnAsnArgProGlyPheYsLeuGlnYrGlnAlaYrGlyLeuGlnGln 1839
Db 6339 GACCATTCGCAAAACCGGCAAGGATTTAAACTTGTCTTACAAAGCTTATGAATTAAGAAC 6398
Qy 1840 CysProAspProGlnProPheAlaAsnGlyIleValArgGlyAlaGlyYrAsnValGly 1859
Db 6399 TGTCCAGATCCACCCCATTTCAAGATGGTACATGATCAACTCGGATTTACAGCTGGGG 6458
Qy 1860 GlnSerValThrPheGlnCysLeuProGlyYrGlnLeuThrGlnHisProValLeuThr 1879
Db 6459 CAATCAGATCTTCTGAGTGTATCTGTGGATCATTTTAATAGCCATCTGTCTCACT 6518
Qy 1880 CysGlnHisGlyThrAsnArgAsnYrAspHisProLeuProYsCysGlyValProCys 1899
Db 6519 TGTTCAGCATGGGATCAACAGAACTGAACTACCTTTTCCAAAGTGTGATGCCCTTGT 6578
Qy 1900 GlyIleAsnIleThrSerSerAsnGlyThrValYrSerProGlyPheProSerProYr 1919
Db 6579 GGGTCAACAGTAACTTCTCAGAACGCGCACATCTACTCCCTGGCTTCTCATGATGAT 6638
Qy 1920 SerSerSerGlnAspCysValTrpLeuIleThrValProIleGlyHisGlyValArgLeu 1939
Db 6639 CGGATCTCGAAGAGCTGATTTGGCTCATACGGTCTCCAGGGCAGAGATTTATCATC 6698
Qy 1940 AsnLeuSerLeuLeuGlnThrGlnProSerGlyAspPheIleThrIleTrpAspGlyPro 1959
Db 6699 AACTTCACCTGTTTACAGACGGAAGCTGTCAACGATTTACATTCGCTTTGGGACGCTCC 6758
Qy 1960 GlnGlnThrAlaProArgLeuGlyValPheThrArgSerMetAlaYsLeuThrValGln 1979
Db 6759 GATCAGAACTCAACCCAGCTGGAGTTTCAGTGGCAACAGGCCCTGAAACGCGCTAT 6818
Qy 1980 SerSerSerAsnGlnValLeuLeuYsPheHisArgAspAlaAlaThrGlyGlyIlePhe 1999
Db 6819 AGCTCCACCAACCAAGTCTCTCTCAAGTTCCACAGGCACTTTTCAATAGAGGCTTCTT 6878
Qy 2000 AlaIleAlaPheSerAlaYr-----ProLeuThrYs----- 2010
Db 6879 GTCTCAATTTCCACGGCACTGATTTTCACTCGGATTAAGATGACGAGAATTCATG 6938
Qy 2011 TGGTGTTCACGAGTGTGCCAGCGCTTGTTCAGCTGAAGTTTCTTATTCAGCC 6998
Db 2021 GluValValThrGluAsn 2026
Db 6999 GAGGCGGTGATGATTC 7016

```

RESULT 5

US-10-276-934-6

; Sequence 6, Application US/10276934

; Publication No. US20030180750A1

; GENERAL INFORMATION:

; APPLICANT: University of Leeds

; APPLICANT: Markham, Alexander F.

; APPLICANT: Jackson, Andrew P.